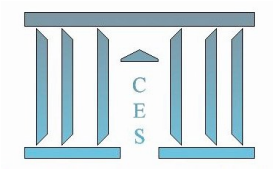
A fully non-parametric heteroskedastic model

Matthieu Garcin, Clément Goulet

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers



A fully non-parametric heteroskedastic model

Matthieu GARCIN, Clément GOULET

2015.86

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://centredeconomiesorbonne.univ-paris1.fr/
ISSN : 1955-611X

A fully non-parametric heteroskedastic model

Matthieu Garcin*, Clément Goulet†

September 30, 2015

Abstract

In this paper we propose a new model for estimating returns and volatility. Our approach is based both on the wavelet denoising technique and on the variational theory. We assess that the volatility can be expressed as a non-parametric functional form of past returns. Therefore, we are able to forecast both returns and volatility and to build confidence intervals for predicted returns. Our technique outperforms classical time series theory. Our model does not require the stationarity of the observed log-returns, it preserves the volatility stylised facts and it is based on a fully non-parametric form. This non-parametric form is obtained thanks to the multiplicative noise theory. To our knowledge, this is the first time that such a method is used for financial modelling. We propose an application to intraday and daily financial data.

1 Introduction

Various models have been used for estimating financial market volatility. The stake behind such estimations is to take into account volatility stylized facts with as few assumptions on the observable returns as possible. Insofar as that volatility provides a measurement of an asset return variability over a time period, at least three stylised facts have to be considered: leverage effect, clustering and negative asymmetry. Leverage effect can be defined as the negative correlation between asset returns and volatility. Clustering implies that large (resp. small) price moves tend to be followed by large (resp. small) ones. Negative asymmetry implies a left-skewed distribution of financial returns. Mishra, Su and Ullah provide a survey on the different classes of time-varying volatility models [24]. Two are commonly used: the conditional variance, which is time-discrete, and the stochastic volatility, which is time-continuous.

On the one hand, the first class of model developed in the framework of conditional variance is the autoregressive conditional heteroscedasticity (ARCH) model [10]. This model allows the volatility to be a linear function of past log-returns. As it is well known that the relationship between volatility and past log-returns is non-linear, Bollerslev has extended this class of model to the case where the volatility is a non-linear function of past log-returns and past volatility [3]. Both classical ARCH and GARCH models are parametric, they reproduce volatility clustering effect, they assume a probability law of the observed returns and finally they require stationarity of the observed returns. These models have at least four limits:

- ▷ First, the classical ARCH and GARCH models suffer from a symmetry around zero and a weak kurtosis, whereas the log-returns distributions of financial assets are leptokurtik and

*Natixis Asset Management-LabEx REFI
†University Paris 1-LabEx REFI

asymmetric. Indeed, ARCH and GARCH models assume that the volatility is independent of the sign of the last log-return even though observations showed that the volatility increase due to a negative log-return is generally higher than the volatility increase induced by a positive one. Some GARCH extensions, such as GARCH-GJR [16] or E-GARCH [25], take into account asymmetry.

- ▷ Second, they rely on a stationarity assumption of log-returns. Two main extensions including non-stationarity for GARCH models have been provided: the local stationary multiplicative volatility model [11] and local stationary varying GARCH coefficients [6].
- ▷ Third, GARCH estimators suffer from some drawbacks regarding their convergence. Winker and Maringer compared the two main approaches for the estimation of GARCH parameters: the maximum-likelihood estimator and the threshold-accepting estimator [30]. They noticed that deterministic algorithms do not provide convergent estimators and they recommended stochastic algorithms, even if these stochastic algorithms introduce a measurement noise.
- ▷ Fourth, the latest developments of GARCH models are either based on more complicated probability distributions of the residuals or on semi-parametric approaches [4, 7, 29]. They are often complex to estimate and do not usually much better fit observable returns nor volatility. Consequently, these models are not widely used by practitioners. The model we propose is fully non-parametric, it is based on innovative mathematical tools and it is quite easy to implement.

On the other hand, stochastic volatility models such as Heston model [19] or Stochastic Alpha Beta Rho (SABR) model [18] are based on stochastic calculus theory. These models are time-continuous and are widely used for estimating the spot volatility through the implicit volatility. Mikhailov and Nogel pointed out some drawbacks of these models [23]. First, these models are parametric. Second, they are time-continuous whereas the observed processes are time-discrete. Finally the estimates obtained are parameter-sensitive. Some extensions of volatility continuous models include semi-parametric approaches. For example, Kanaya and Kristensen proposed a non-parametric estimate for instantaneous volatility in the framework of stochastic volatility [20]. The estimate they designed is based both on the filtering theory and on kernel techniques. However, this approach requires stationarity for the observable process and the estimate of volatility is only built for high-frequency data.

The model we propose follows on form the improves of time series models and has the following properties: non-stationarity is allowed (including jumps), it preserves the volatility stylised facts posited above and it is fully non-parametric. We assess that the instantaneous volatility can be expressed as a non-parametric form of past returns. The expected volatility is the consequence of decisions made by the market actors, such as market-makers, traders, arbitrage strategies, between time t and $t+1$. We can consider the volatility as a reaction to observed past returns. To estimate this reaction function we use an iterative algorithm based both on the wavelet theory and on variational calculus. The problem we solve of a non-parametric and time-varying form of the volatility is close to a denoising problem in the framework of an affine noise and therefore it is close to the multiplicative noise theory. Indeed, we assume that the observed returns are decomposed into a time-varying average term and the product of a volatility term by a noise represented by a unit Gaussian random variable. The time-varying average term is obtained by using the denoisement of the empirical wavelet coefficients obtained by a projection and filtering of observed returns on a wavelet basis. The volatility functional form is estimated using a variational approach conditioned to the law of the noise.

The contributions of our paper are twofold. First, we propose an innovative and performing approach for modelling financial log-returns and volatility, without any stationarity assumption on

the observed system. Second, we provide an extension of a multiplicative noise removal method, based on a variational approach, to the case of real-support Gaussian noise.

Our article is organised as follows. In Section 2, we present the estimation algorithm. In Section 3, we test its convergence on simulated data. In Section 4, we apply it on financial data and compare it to some GARCH-oriented models. The results obtained outperform classical time series models.

2 Estimation algorithm

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

$$\left\{ \begin{array}{rcl} y(t) & = & x(t) + g(\mathcal{Y}(t))\varepsilon_t \\ \mathcal{Y}(t) & = & (y(t-1), ..., y(t-h))', \end{array} \right. \tag{1}$$

in which y is the observed price return of an asset, x is an unknown function corresponding to the return of the fundamental asset value, $h \geq 1$ is an integer indicating the number of lags in the information, $\mathcal{Y}$ is the lagged information of price returns and $g(\mathcal{Y}(t))\varepsilon_t$ is the noisy part of the observed price at time t. More precisely, $\varepsilon_1, ..., \varepsilon_T$ are independent unit Gaussian random variables and g is an unknown and positive function. In this model, one assumes that the standard deviation of the noise is a function of the observed price returns. Implicitly, it means that the slippage amplitude g is a consequence of decisions made by investors and market makers. These decisions are based on observations of successive price variations, from $t - h - 1$ to $t - 1$.

Before we use this model for predicting future price returns with confidence intervals as well as defining trading strategies, we first have to estimate the model. The challenge hence consists in the estimation of both x and g. For this purpose, we will use both wavelet and variational methods. Indeed, x is an unknown function representing the fundamental price return. It can therefore present some inhomogeneities, contrarily to g, which smoothly links price return observations and decisions made by investors and market makers. On the one hand, wavelet denoising methods, since they consist in the analysis of a signal into localised elements, allow a good accuracy for inhomogeneous functions. They are thus well suited to x. On the other hand, we can impose smoothness constraints in variational methods. We are then inclined to estimate g with a variational approach.

In the next subsections, we present the algorithm proposed to estimate x and g.

2.1 Preliminary steps of the algorithm

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

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

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

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

k, by:

$$\langle y, \psi_{j,k} \rangle = \sum_{t=0}^{T} y(t)\psi_{j,k}(t). \tag{2}$$

Further details on wavelets and its use to denoise time series can be found in [21, 13].

The second preliminary step of the algorithm relates to the variational approach. In this method, we will minimize an integral of a function in which g appears. When g is multidimensional, that is when $h > 1$, we can face an empirical multidimensional integral with an irregular grid. Several methods are possibles, such as a distortion of the observation grid, a signal interpolation or Voronoi cells [13]. Due to its simplicity, we choose a distortion of the grid and more precisely we use a line integral. We thus select a bijective function $\theta : \{0, ..., T\} \to \{0, ..., T\}$. θ links the new time variable t to the natural observation time $\theta(t)$. It leads to the path $\mathcal{Y} \circ \theta$ along which the integral of g is empirically calculated. The choice of θ can be done once for all in the preliminary steps of the algorithm. The idea is to minimize the Euclidean distance between $\mathcal{Y}(\theta(t))$ and $\mathcal{Y}(\theta(t+1))$ for all $t \in \{0, ..., T-1\}$. For example, if $h = 1$, we choose θ so that the observed returns are sorted: $\mathcal{Y}(\theta(0)) \leq \mathcal{Y}(\theta(1)) \leq ... \leq \mathcal{Y}(\theta(T))$. For higher h, the choice of θ may be related to the travelling salesman problem, for which an approximation algorithm may be used. Whatever the choice made for θ, there will be an impact on the estimate of g when $h > 1$. Indeed, in our variational problem, we aim to minimize the squared derivative of g over all the observations. But this derivative is a derivative in only one direction while using the line integral, instead of a derivative thought as a gradient. Therefore, when we choose a particular θ we may incidentally favour the smoothness of g at each observation point in one direction and not necessarily in all the directions. However, this limitation does not appear in dimension $h = 1$.

2.2 The algorithm for estimating x and g

We achieve the estimation of x and g iteratively:

1. We begin by initializing the series of estimators: $g_0 = M/0.6745$, where M is the median of the absolute value of the wavelet coefficients of y at the finer scale, as usually done for wavelet denoising techniques with an homogeneous variance of the noise [8]. Indeed, $M/0.6745$ is a robust estimator for the noise standard deviation.

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

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

   for a level-dependent threshold $\Lambda_{i,j,k} = \lambda_i \sqrt{\langle (g_i \circ \mathcal{Y})^2, \psi_{j,k}^2 \rangle}$ where λ_i is a parameter. Examples indeed show that a level-dependent threshold performs much better than a constant threshold [15]. The choice for λ_i may be arbitrary, but we prefer to optimize it, that is to choose the value of λ_i which minimizes an estimate of the reconstruction error. This is the aim of *SureShrink* [27, 9, 21]. The estimate of the reconstruction error is

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

where

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

because $\langle (g_i \circ \mathcal{Y})^2, \psi_{j,k}^2 \rangle$ is the variance of the empirical wavelet coefficient $\langle y, \psi_{j,k} \rangle$ [14]. Conditionally to y, $\bar{\mathcal{S}}_i$ is an unbiased estimate of the reconstruction error. Any basic optimization algorithm allows then to get the λ_i minimizing $\bar{\mathcal{S}}_i$. Thus, $\Lambda_{i,j,k}$ and $F_{i,j,k}$ for all j and k are now defined. We can hence write the estimate x_i of the function x as:

$$x_i(t) = \sum_j \sum_k F_{i,j,k}(\langle y, \psi_{j,k} \rangle) \psi_{j,k}(t),$$

for each $t \in \{0, ..., T\}$.

3. We now use the i-th estimate of x to estimate g. This is similar to estimating a signal disrupted by a multiplicative noise. We can then use a variational approach to estimate g. In the literature devoted to multiplicative noise, the case of a Gaussian variable is often excluded since the noisy signal is positive. However, in our case, the estimate of g, which stems from the estimate x_i, is evaluated from the noisy signal $y - x_i$, which is not expected to be positive at each t. The idea of the variational method is to find a function g_{i+1} which will be the solution of an optimization problem. This optimization problem consists, for each observation time, in maximizing the likelihood of $y - x_i$ conditionally to g_{i+1} given that the noise is a Gaussian noise. In addition to that local criterion, we add a global constraint. This constraint is a penalty term which favours the smoothness of g_{i+1}. Our method is partially inspired by the one proposed by Aubert and Aujol for removing Gamma multiplicative noise [2]. It leads to the following equation for the estimate $g_{i+1} \circ \mathcal{Y}$ of $g \circ \mathcal{Y}$, where we introduce $\mathcal{G}_{i+1}$ which we define[1] by $\mathcal{G}_{i+1} = g_{i+1} \circ \mathcal{Y} \circ \theta$:

$$\mu \frac{(\mathcal{G}_{i+1})^2 - (y \circ \theta - x_i \circ \theta)^2}{(\mathcal{G}_{i+1})^3} - \frac{d^2}{dt^2} \mathcal{G}_{i+1} = 0, \tag{3}$$

where $\mu > 0$ is a parameter which allows to tune the priority between smoothness of g_{i+1} and accuracy of the model by means of the maximum-likelihood approach. More precisely, the smoothness of g_{i+1} increases when μ decreases. Details about how this equation is obtained are given in appendix A. Then, in order to solve numerically this equation, we use a dynamical version of it which is expected to lead to a steady state after some iterations of the series of estimators $(\mathcal{G}_{i+1,n})_n$ of $\mathcal{G}_{i+1}$:

$$\frac{\mathcal{G}_{i+1,n+1} - \mathcal{G}_{i+1,n}}{\delta} = \frac{d^2}{dt^2} \mathcal{G}_{i+1,n} - \mu \frac{(\mathcal{G}_{i+1,n})^2 - (y \circ \theta - x_i \circ \theta)^2}{(\mathcal{G}_{i+1,n})^3},$$

where δ is a parameter controlling the speed to which $(\mathcal{G}_{i+1,n})_n$ evolves. More precisely, for each $t \in \{0, ..., T\}$, the series $(\mathcal{G}_{i+1,n}(t))_n$ is iteratively defined by:

$$\begin{cases} \mathcal{G}_{i+1,0}(t) & = & Median\{|y(s) - x_i(s)|\}/0.6745 \\ \mathcal{G}_{i+1,n+1}(t) & = & \mathcal{G}_{i+1,n}(t) + \delta \left[\mathcal{G}_{i+1,n}(t+1) - 2\mathcal{G}_{i+1,n}(t) + \mathcal{G}_{i+1,n}(t-1) - \mu \frac{\mathcal{G}_{i+1,n}(t)^2 - (y(\theta(t)) - x_i(\theta(t)))^2}{\mathcal{G}_{i+1,n}(t)^3} \right]. \end{cases}$$

$\mathcal{G}_{i+1,n}$ is expected to converge towards $\mathcal{G}_{i+1}$ when n tends towards infinity. The convergence is sensitive to the choice of parameters. In particular, the higher δ, the faster the initial estimator $\mathcal{G}_{i+1,0}$ of $\mathcal{G}_{i+1}$ will be distorted. However, if δ is too big, fine adjustments from $\mathcal{G}_{i+1,n}$ to $\mathcal{G}_{i+1,n+1}$ will often be excluded and the convergence towards a steady state will be compromised.

[1] $\mathcal{G}_{i+1}$ will be more clearly defined if we write its domain and codomain: $\mathcal{G}_{i+1} : \{0, ..., T\} \to \mathbb{R}$, since it is obtained by the composition of $\theta : \{0, ..., T\} \to \{0, ..., T\}$ with $\mathcal{Y} : \{0, ..., T\} \to \mathbb{R}^h$ and $g_{i+1} : \mathbb{R}^h \to \mathbb{R}$.

3 Data simulation and algorithm application

In this section, we study the convergence of the variational part of the algorithm when the functional form g is known *a priori*. We recall that the main contribution of this paper is to provide a Fully Non-Parametric Heteroscedastic (FNPH) form for volatility modelling. Hence, we reduce the model introduced in Section 2 into a classical multiplicative noise problem. For sake of simplicity, the past observations do not influence the non-parametric functional form g in this section. Since g is a function of the time only, we estimate g_i by the series $g_{i,n}$, where $g_{i,n} = \mathcal{G}_{i,n}$, for $n \in \mathbb{N}$. Moreover, in absence of the average return component x, we do not need to estimate x and g iteratively. Therefore i is fixed. In this section, the studied model is then defined as follows:

$$y(t) = g(t)\varepsilon_t, \quad t \in \{0, ..., T\}.$$

To study the convergence of the estimated sequence $(g_n)_{n\in\mathbb{N}}$ toward $\hat{g}^2$, given (μ, δ), we will minimize two distances: the Hellinger distance and the Euclidean distance. We recall that δ controls the speed of convergence of g_n toward $\hat{g}$ when n tends toward infinity. As we previously said, the main issue raised by the choice of δ, is that if δ is too small, the rate of convergence will be small. In addition, at the steady state, g will be misestimated. On the contrary, if δ is too large, the estimation of g will take into account the spikes originated by the noise term, and so $\hat{g}$ will diverge. In other words, the empirical noise obtained at the steady state will differ significantly from the theoretical law of the noise. Hence, to obtain the convergence of g_n toward g when n goes to infinity, we choose δ that minimizes a distance between the theoretical law of the noise and the empirical law of the noise.

As in our model both the theoretical law of the noise and its moments are known, we choose δ by minimizing the Hellinger distance[3] between the empirical noise, $\hat{\varepsilon}_t = \frac{y(t)}{\hat{g}(t)}$, and the theoretical noise ε_t. At this first step, we point out that μ is fixed arbitrarily *ex ante*. Once we have obtained the minimizer δ^*, we can select the smoothing parameter μ. To do so, we minimize the Euclidean distance between g_n and g [4]. If μ is too large, then the variational term $\frac{g_{n-1}^2 - y^2}{g_{n-1}^3}$ will bring the roughness of the observed signal. If μ is too weak, then the estimated $\hat{g}$ will minimize the observed variations of g. So, we choose μ that minimizes the Euclidean distance between g and $\hat{g}$.

In Figure 1, we simulate y with a non-linear and time-dependent g and so the process $g(t)\varepsilon_t$ is non-stationary. The objective of dealing with non-stationary processes is to get closer to the properties of observable processes such as log-returns of financial assets.

In Figure 2, we represent the convergence of the Hellinger distance calculated between the theoretical probability density function of the noise and the empirical probability density function of the noise, when n goes to infinity. We observe that, even if the behaviour of the distance is locally erratic, we obtain the convergence in distribution of the estimated probability density function of the empirical noise toward the probability density function of the theoretical noise. The choice of δ^* was done for $n = 1000$ and $\mu = 1$. For this value, the obtained Hellinger distance is steady. Then, we choose the optimal smoothing parameter μ by minimizing the Euclidean distance between the true functional form g and our estimate $\hat{g}$. We are aware that, when the functional form g is not known *a priori*, the smoothing parameter must be fixed arbitrarily in order to satisfy another criterion[5]. Here, g is smooth, time-dependent and non-linear. On subfigure (a), for n bigger than

[2] We set $\lim_{n\to\infty}(g_n) = \hat{g}$. $\hat{g}$ is an estimate of g.

[3] The Hellinger distance between two probability density functions P and $\hat{P}$ taking value on the same support is defined by $\mathcal{H}^2(dP, d\hat{P}) = \int_{\mathbb{R}} \left(\sqrt{dP(x)} - \sqrt{d\hat{P}(x)}\right)^2 dx$.

[4] This methods is an oracle because it uses the true form g.

[5] In Section 4, this criterion is the closeness between the scale of g and the scale of an empirical instantaneous volatility

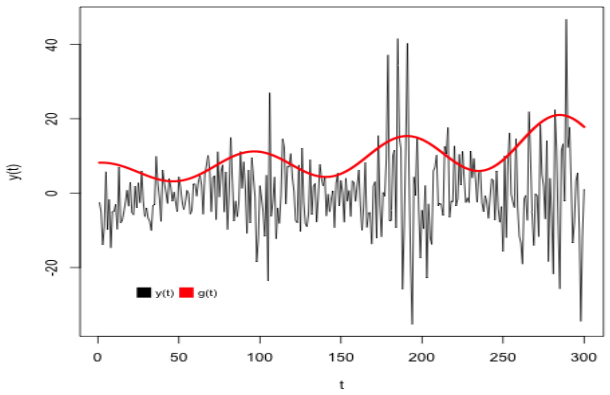


Figure 1: Simulation of $y(t) = g(t)\varepsilon_t$.
We simulate, for $T = 300$, $y(t) = g(t)\varepsilon_t$. We set $g(t) = \left(cos(\frac{t}{15}) + 2\right) e^{\frac{t}{300}+1}$. g was chosen in order to have a non-stationary process $g(t)\varepsilon_t$. In fact, we test whether the algorithm is able to capture the non-stationarity induced by g. The multiplicative noise term is a Gaussian white noise with variance equal to 1.

1000, the Hellinger distance is steady and its value is quite low[6]. On subfigure (b), we observe that g_n and g are increasingly close as n grows.

In Figure 3, we plot our estimate $\hat{g}$. Such estimate captures the non-linearity property of g as well as its time-dependency. As $\left(\frac{y(t)}{g_n(t)}\right)$ pass a normality test[7], the multiplicative noise estimate is not far from the theoretical distribution of the multiplicative noise. In conclusion, the algorithm we developed fulfils in recovering a non-linear and a time-dependent function g. In the next section, we apply the algorithm on financial data.

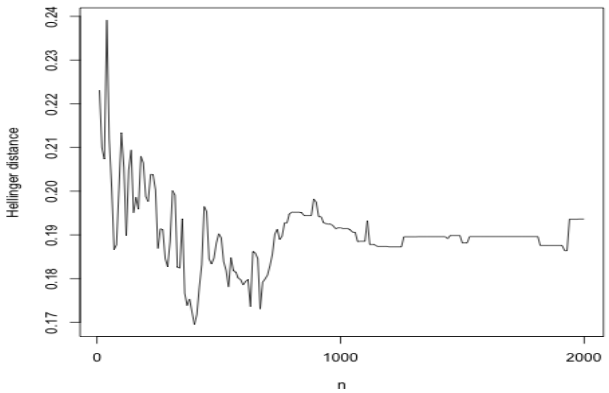
4 Empirical application

In this section we apply the model on intraday and daily financial data. As intraday data, we selected three exchange rates: EUR/USD, EUR/GBP and GBP/USD and we estimated the model during the month of February 2015 over 15-minute intervals[8]. As daily data, we selected the French index CAC 40 between 1st March 1990 and 16th June 2015. First, we present the results of the algorithm for the estimation of $g(\mathcal{Y}(t))$. Then, we forecast the instantaneous volatility and we compare the obtained results with other conditional volatility models (Table 1). Finally, we propose an application based on a simulated trading strategy.

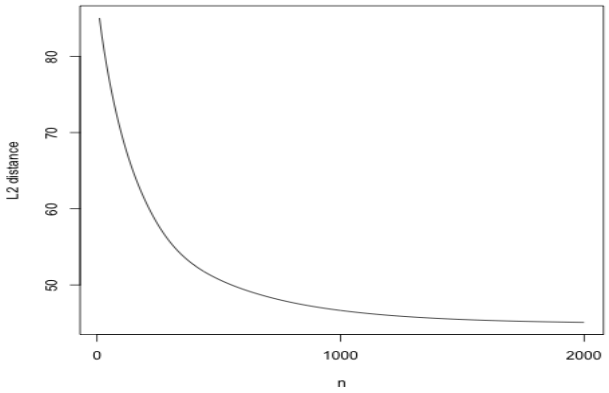
[6] How low can be the Hellinger distance? We compare such a distance between the $\mathcal{N}(0,1)$-probability distribution and its empirical version obtained with 1000 $\mathcal{N}(0,1)$-random variables. The Hellinger distance should be at the lowest. We repeat the operation until we get 1000 Hellinger distances corresponding to 1000 different sets of 1000 $\mathcal{N}(0,1)$-random variables. Then, the average Hellinger distance is equal to 0.11. This value isn't nothing compared to the Hellinger distance computed between the distribution of $\frac{y(t)}{g_n(t)}$ and the $\mathcal{N}(0,1)$-probability distribution, which is 0.19.

[7] The normality test is Jarque Bera. The p-value=0.62 and $\chi^2 = 0.92$.

[8] The tick-by-tick data are available on *Truefx.com*. For the three exchange rates we computed the 15-minute prices by calculating the average price among all tick prices.



(a) Hellinger distance



(b) Euclidean distance

Figure 2: Convergences of g_n for the Hellinger distance and the Euclidean distance.
After having minimized the Hellinger distance between ε_t and $\hat{\varepsilon}_t$ for $n = 1000$ and $\mu = 1$, we found that $\delta^* = 0.24$. Then, for $n = 1000$ and δ^*, we minimized the Euclidean distance between g and g_n, we found that $\mu^* = 3.41$. Subfigure (a) represents the behaviour of the Hellinger distance between $\hat{\varepsilon}_t$ and ε_t when n goes to infinity. Subfigure (b) shows the evolution of g_n toward g when n goes to infinity.

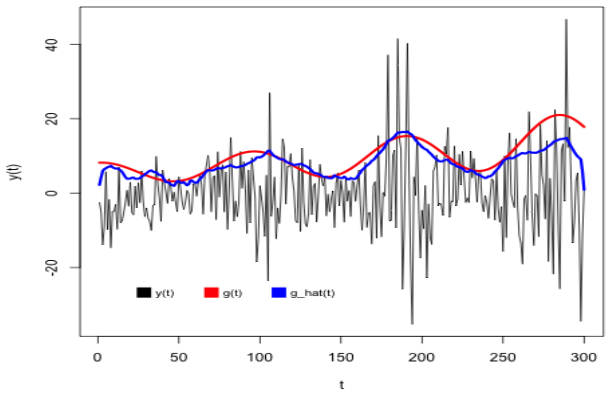


Figure 3: Estimate of g.
After choosing the vector (δ^*, μ^*) that minimizes both Hellinger distance and then the Euclidean distance, we have estimated g by g_n for $n = 1000$.

4.1 Estimation of $g\left(\mathcal{Y}(t)\right)$

Before estimating g, we shortly discuss the wavelet family, the number of vanishing moments and the resolution level to which the wavelet series is truncated. First, the wavelet basis chosen is a Daubechies one, which is orthogonal and compactly supported. The wavelet basis is obtained thanks to the orthonormalization of a Riesz basis. Daubechies wavelets are asymmetric. This restricts the construction of regular wavelets to the real line [21]. Furthermore, the empirical wavelet coefficients obtained are not orthogonal. This affects the signal decomposition. A wavelet has κ vanishing moments if its scaling function can approximate polynomials of degree less or equal to κ. For smooth enough pure signals, which can locally be developed in Taylor series, a particular κ, depending on the values of the successive derivatives, can be indicated so as to decompose it parsimoniously in a wavelet basis. In this paper, no assumption is made on the existence and value of the derivatives of the pure signal. Therefore, we simply fix $\kappa = 4$. Finally, the higher the resolution parameter j chosen, the more inaccurate the projection of the signal on the truncated wavelet family. In finance, a rough projection of log-returns on a wavelet family allows to capture their trend. The trend of the observed log-returns can be seen as an instantaneous average. Thus, if the resolution parameter is too low, it will be hard to distinguish the trend from the noise. In this section, we arbitrarily fix $j = 4$. We show the signal y and its projection x in Figure 4.

The main challenge of the FNPH model is to provide a relevant estimate of $g\left(\mathcal{Y}(t)\right)$. In this paper, we only provide $g(\mathcal{Y}(t))$ for $h = 1$. To calculate $g(\mathcal{Y}(t))$ for $h > 1$, we should approximate the path along which to integrate the likelihood using methods such as nearest neighbours or algorithms related to the travelling salesman problem. We are aware that in practice it is hard to find a causal relation between a price move some ticks behind and a volatility variation right now, but the aim of our non-parametric approach is to capture any past price influence on the instantaneous volatility variation. As the theoretical form $g(\mathcal{Y}(t))$ for an observable system is unknown, we cannot compare graphically the estimate $\hat{g}(\mathcal{Y}(t))$ to $g(\mathcal{Y}(t))$. To overcome this issue, we compare each form $\hat{g}(\mathcal{Y}(t))$ obtained to an estimate of the instantaneous volatility $\hat{\sigma}_t$[9].

The choice of the parameters (δ, μ) is done such that the scale of $\hat{g}(\mathcal{Y}(t))$ is close to the scale of $\hat{\sigma}_t$.

[9] We set $\hat{\sigma}_t = |y(t) - y(t-1)|$.

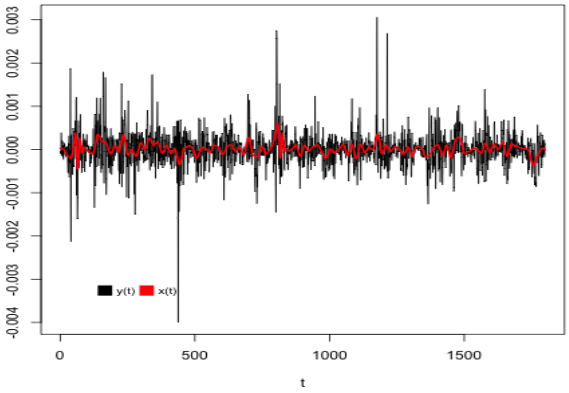


Figure 4: Approximation of $x_0(t)$ for GBP/USD log-returns.
The resolution level is equal to 4 and the number of vanishing moments is equal to 4. $g_0 = 0.0012$.

Figure 5 shows the estimation of $g(\mathcal{Y}(t))$ for GBP/USD log-returns between 1st of February 2015 and the 28th of February 2015 (15-minute observations). The algorithm converges rapidly. For $i = 2$ we obtain a stable estimate $g_2(\mathcal{Y}(t))$. The estimation of the sequence $(\mathcal{G}_{i+1,n}(t))_n$ is sensitive to the number of iterations n. The issues raised by the choice of n are discussed in the previous section. For the four log-returns used, we obtain a smile relationship between the instantaneous log-returns and the one-time-ahead volatility. This relationship captures both leverage effect and volatility clustering. Clustering effect because high returns in absolute value are followed by high volatility. Leverage effect because a downward return creates more volatility than an upward one.

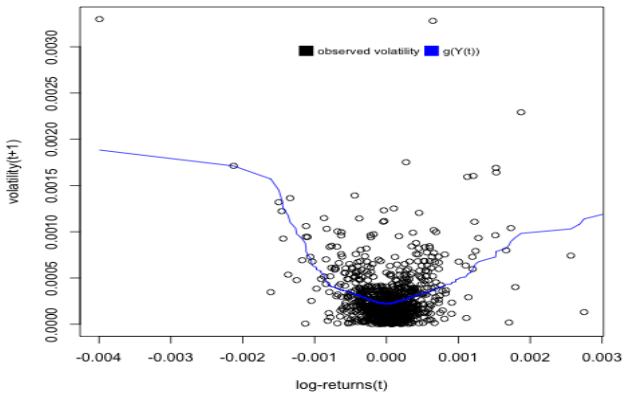


Figure 5: Estimates of g_1 for GBP/USD log-returns.
The bullet points are the instantaneous volatility, the blue line is the estimate of g after one iteration.

4.2 Model validation

To validate the model we designed, we compare it to the classical time series framework. We focus on the same example than in the previous section: GBP/USD log-returns between 1st of February 2015 and the 28th of February 2015 with 15-minute observations. We treat three well-known models whose equations are given in Table 1: the GARCH model [3], the E-GARCH model [25] and the GARCH-GJR model [16]. To improve the goodness of fit of the three GARCH-oriented models, we add an autoregressive term. For each approach, we will compare the distribution of the empirical residuals. The purpose of this study is to find whether our model better matches the statistical properties of financial assets log-returns. Various papers have studied these statistical properties [22, 5, 26]. From now on, we focus on five stylised facts: conditional heteroscedasticity, volatility clustering, leverage effects, fat tails and asymmetric distributions.

Model	**Equation**
GARCH(1,1)	$\begin{cases} y_t = \phi y_{t-1} + \sqrt{h_t^2}\varepsilon_t \\ h_t^2 = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}^2 \end{cases}$
E-GARCH(1,1)	$\begin{cases} y_t = \phi y_{t-1} + \sqrt{h_t^2}\varepsilon_t \\ log(h_t^2) = \omega + \alpha\left(\lvert\varepsilon_{t-1}\rvert - \mathbb{E}\left[\lvert\varepsilon_{t-1}\rvert\right]\right) + \gamma log(\varepsilon_{t-1}^2) + \beta log(h_{t-1}^2) \end{cases}$
GARCH-GJR(1,1)	$\begin{cases} y_t = \phi y_{t-1} + \sqrt{h_t^2}\varepsilon_t \\ h_t^2 = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}^2 + \theta \mathbb{1}_{y_{t-1}<0} y_{t-1}^2 \end{cases}$

Table 1: Time series models.
All the innovation processes are assumed to be Gaussian.

First, we estimate the four models on the first 1500 observations and we study the distribution of the residuals obtained. Second, we forecast the log-returns and the volatility with a rolling window of 400 observations for the four models[10]. Finally we study the forecast errors of volatility and the forecast errors of log-returns.

4.2.1 Distribution of the residuals

To backtest FNPH model, given by equation 1, we first estimate x_1 and g_1 and then we compute the residuals $\hat{\varepsilon}_t$ given by:

$$\hat{\varepsilon}_t = \frac{y(t) - x_1(t)}{g_1(\mathcal{Y}(t))}, \quad t \in \{0, ..., T\}.$$

Figure 6 shows the empirical log-distribution of the residuals. After having tested their normality, we conclude that the residuals obtained by the FNPH model are Gaussian (see Table 3). In other words, as the law of the residuals is the same as the law of the theoretical residuals, we can say that the FNPH model is well specified and reproduces the stylised facts of the log-returns.

Then, we fit GARCH(1,1), E-GARCH(1,1) and GARCH-GJR(1,1) on the first 1500 GBP/USD observations[11]. Before estimating the parameters of the three models, we tested the stationary assumption of the observed log-returns[12]. This assumption is not verified. We conclude that the estimates of the parameters may be biased. The results of the estimation are available in Table 2. The FNPH model has clearly an higher likelihood than the GARCH oriented models.

[10] The results for the three other sets of log-returns are available in Appendix B.
[11] The three GARCH models are estimated by the quasi-maximum likelihood technique.
[12] The test of stationarity has being done by the unit-root methodology. We found p-values < 0.01.

Nevertheless, asymmetric GARCH-oriented models (E-GARCH and GJR-GARCH) better fit the USD/GBP log-returns than the standard symmetric GARCH model.

Model	ϕ	ω	α	β	θ	γ	$\mathcal{L}$
GARCH(1,1)	$1.75e-1$	$4.70e-9$	$6.71e-2$	$8.98e-1$	-	-	9683
E-GARCH(1,1)	$1.78e-1$	-1.45	$1.23e-2$	$9.05e-1$	-	$4.75e-1$	9774
GJR-GARCH(1,1)	$1.90e-1$	$4.20e-9$	$5.80e-2$	$8.68e-1$	$1.12e-1$	-	9707
FNPH	-	-	-	-	-	-	10382

Table 2: Estimation results of the four processes.
The processes are fitted on the 1500 first observations of the GBP/USD log-returns.

Furthermore, the empirical residuals of the four models have been computed (see Table 3). The residuals of the FNPH model are less skewed than the residuals of GARCH-oriented models. In other words, the residuals of GARCH-oriented models are asymmetric and so these models underestimate/overestimate log-returns. In addition, the kurtosis in GARCH models are bigger, which implies that there is a frequent number of extreme deviations. After having tested their distributions, we conclude that their normality is rejected. Figure 6 shows the log-distributions of the empirical residuals for the four models. As shown by their moments, GARCH-oriented residuals have bigger tails than FNPH residuals. It highlights that GARCH-oriented models are misspecified and depict therefore less accurately extreme events than the FNPH model.

Model	Expectation	Skewness	Kurtosis	P-value of normality test
GARCH(1,1)	$2.82e-2$	$3.45e-1$	11.90	$3.26e-5$
E-GARCH(1,1)	$3.07e-2$	$4.95e-1$	8.65	$1.08e-4$
GARCH-GJR(1,1)	$2.57e-2$	$5.58e-1$	11.47	$2.37e-4$
FNPH	$1.31e-3$	$2.43e-2$	-0.70	$1.52e-1$
$\mathcal{N}(0,1)$	0	0	3	1

Table 3: Study of empirical residuals for the four models.
After having estimated models on the 1500 first observations, the empirical residuals are computed. The residuals of the four models are Gaussian according to the assumption. The normality is tested by the Kolmogorov-Smirnov procedure.

4.2.2 Volatility forecast

To compare the forecast ability of the FNPH model to the GARCH(1,1), E-GARCH(1,1) and GARCH-GJR(1,1) models, we first study volatility forecasting errors. As we previously said, the forecasts are done recursively with a 400-day rolling window. GARCH models are re-estimated each day. We point out that g doesn't need to be re-estimated. Indeed, we assess that the relationship between log-returns in t and volatility in $t+1$ is structural. So, to have a relevant estimate of g, the dataset should have turmoil periods and calm periods. As this condition is fulfilled by the dataset USD/GBP, we do not re-estimate g at each time step.

For GARCH-oriented models, we forecast the next-period volatility of returns using the square root of its conditional variance.

Once we have forecasted the volatility during 291 steps, we study their errors using the Mean Absolute Deviation (MAD) and the Root Mean Square Error (RMSE, the standard deviation of

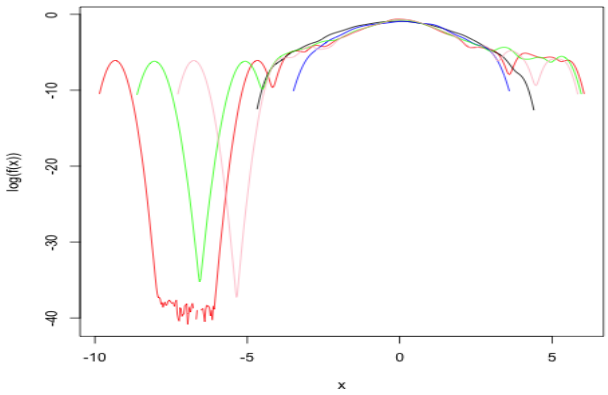


Figure 6: Log-distribution of the residuals for the four models compared to the log-distribution of $\mathcal{N}(0,1)$.
After having estimated models on the 1500 first observations, the empirical residuals are computed. We observe that the FNPH residuals density (blue) is much less skewed that the others. GARCH(1,1) log-density is red. E-GARCH(1,1) log-density is pink. GARCH-GJR(1,1) log-density is green. $\mathcal{N}(0,1)$ log-density is black.

errors)[13]. Both indicators, by taking account of the average absolute deviation and the average dispersion of errors, permit to compare different models. Table 4 gathers these indicators. Errors obtained for GARCH-oriented models are more skewed and have bigger extreme values than errors obtained for the FNPH model.

	Model	MAD	RMSE
Volatility	GARCH(1,1)	$1.97e-4$	$2.66e-4$
	E-GARCH(1,1)	$1.99e-4$	$2.66e-4$
	GARCH-GJR(1,1)	$1.99e-4$	$2.67e-4$
	FNPH	$1.77e-4$	$2.49e-4$
Log-returns	GARCH(1,1)	$2.38e-4$	$3.27e-4$
	E-GARCH(1,1)	$2.39e-4$	$3.28e-4$
	GARCH-GJR(1,1)	$2.37e-4$	$3.28e-4$
	FNPH	$2.34e-4$	$3.24e-4$

Table 4: Forecast errors for the four models.

We notice the volatility forecasts of the FNPH model outperform the volatility forecasts of GARCH(1,1), E-GARCH(1,1) and GARCH-GJR(1,1). Indeed, the error indicators obtained with the FNPH model are much weaker than the ones obtained by the GARCH-oriented models.

[13] Error indicators are computed between estimates of instantaneous volatility, introduced previously, and volatility forecasts.

4.2.3 Log-returns forecast

To forecast the log-returns by the FNPH model, we have to propose an estimate of $x(t+1)$ conditionally to the model estimated in t. We can rewrite the FNPH model for forecasts:

$$\left\{ \begin{array}{rcl} y(t+1) & = & m(x(t)) + g(\mathcal{Y}(t+1))\varepsilon_{t+1} \\ \mathcal{Y}(t+1) & = & (y(t), ..., y(t-h))', \end{array} \right.$$

where m can either be a parametric or a non-parametric form.

Then, we can compute an estimate of $y(t+1)$ conditionally to the information set $\mathcal{I}_t := \{y(1), ..., y(t)\}$. For each model, the forecasted log-return is defined as follows:

$$\left\{ \begin{array}{ll} \mathbb{E}\left[y(t+1)|\mathcal{I}_t\right] = \phi y(t) & GARCH(1,1) \\ \mathbb{E}\left[y(t+1)|\mathcal{I}_t\right] = \phi y(t) & E-GARCH(1,1) \\ \mathbb{E}\left[y(t+1)|\mathcal{I}_t\right] = \phi y(t) & GARCH-GJR(1,1) \\ \mathbb{E}\left[y(t+1)|\mathcal{I}_t\right] = m(x(t))) & FNPH. \end{array} \right.$$

x is estimated at each time step by the wavelet denoising method introduced in Section 2.

For sake of simplicity, we assess that m follows a classical generalized additive model [17] with a Gaussian kernel. With this assumption, we obtained fully-satisfying results, gathered in Table 4. We could extend m to another non-parametric approach, such as a chaotic approach. Indeed, if the observed system has a negative Lyapunov exponent and if it is characterized by a strange attractor, then we could use a chaotic evolution function to model x [13]. This improvement could be done in further research. To compare each forecast method, we compute the same indicators than those used above. Table 4 presents the study of the forecast error series. We observe that the forecasts done by the FNPH model outperform the forecasts done by GARCH(1,1), E-GARCH(1,1) and GARCH-GJR(1,1). Indeed, the average absolute distance between forecasted returns and realized returns is smaller than those obtained by the GARCH-oriented models. This is mainly due to the better ability of forecasting directly on the denoised system, rather than on the observed log-returns. Again, RMSE for the FNPH model is lower than those obtained by other models. In other words, the error dispersion of FNPH model is weaker than the error dispersion of GARCH-oriented models. This is because the process x, as an approximation of the instantaneous average of log-returns, follows these log-returns in trend, and so it decreases the variation of errors around y.

4.2.4 Density forecast analysis

The four models provide an estimate of the future probability distribution of returns. To compare them, we build a statistic, $\mathcal{W}_{ij} \in \mathbb{R}$, based on the weighted likelihood ratio between a model i and a model j [1]. The principle of this method consists in studying the location of the realized return in the forecasted density. If it is located in the density center, then the model forecasting the first and second moments has a high likelihood. Reversely, if the realized return is located in the density tails, then the model has a low likelihood and thus bad forecast abilities.

Let f_i be the forecasted density obtained by a model i conditionally to the information set $\mathcal{I}_t$ and to the finite-dimension parameter space $\Theta_{i,t}$. For example, if the model i is an AR(1)-GARCH(1,1), then $\Theta_{i,t}$ is a 4-dimensional space. If a model is non-parametric then its parameter set is the empty-set.

After having forecasted two models i and j during T periods, two sequences of forecasted density of realized returns are obtained:

$$\{f_i(y(t+1)|\mathcal{I}_t, \Theta_{i,t})\}_{t=1}^{T} \text{ and } \{f_j(y(t+1)|\mathcal{I}_t, \Theta_{j,t})\}_{t=1}^{T},$$

where $y(t+1)$ is the realized return occurring in time $t+1$. The statistic $\mathcal{W}_{ij}$ is computed as:

$$\mathcal{W}_{ij} = \frac{1}{T}\left(\sum_{t=1}^{T} w(y(t+1)) \left[\log f_i(y(t+1)|\mathcal{I}_t, \Theta_{i,t}) - \log f_j(y(t+1)|\mathcal{I}_t, \Theta_{j,t})\right]\right),$$

where $w(y(t+1))$ is a weight function based on the unconditional distribution of the log-returns[14].

If $\mathcal{W}_{ij} > 0$ then the model i has better forecast abilities than model j and reversely. In Table 5, we observe that the FNPH model has better forecast abilities than the three GARCH-oriented models. Indeed, the weighted likelihood ratio is strictly positive. In other words, realized returns are in average closer to the FNPH forecasted density center than to the GARCH forcasted density centres. Moreover, the weighted likelihood ratio indicates a substantial gain of the FNPH model over the three tested models, compared with the narrow differences between these three parametric GARCH models. It highlights that the fully non-parametric approach can contribute more decisively to accuracy than parametric improvements of the GARCH model.

Model	GARCH(1,1)	E-GARCH(1,1)	GARCH-GJR(1,1)
FNPH	74.80	82.21	78.76

Table 5: Weighted likelihood ratio between the FNPH model an thre other models: GARCH(1,1), E-GARCH(1,1) and GARCH-GJR(1,1). The statistic is computed on the 291 density forecasts of GBP/USD log-returns.

4.2.5 Robustness analysis

Like every model, the estimation and the use of the FNPH model is restricted by the sample size. This is due to the robustness of the noise standard deviation estimate. Indeed, by the central limit theorem, we know that, in order to have a robust estimate of the noise standard deviation, the number of observations should at least be greater than 15 [28]. So the sample size should verify:

$$\frac{T}{2^{j-1}} \geq 15,$$

where T is the sample size and j is the resolution dimension.

For example, if we want to denoise an observable process at a resolution level $j = 4$, then we should have at least 120 observations. Once the sample size verifies the above condition, then the estimation of g can be improved using interpolation techniques.

On the contrary, GARCH-oriented models suffer from an important sensitivity to the number of observations. If the number of observations is below 300, then the amount of available information is not enough for having convergent parameter estimates. But, if the number of observations is too big, then the stationary assumption may be not verified.

[14] Here we assess that it is equal to the normal density function, whose first and second moments are the two first moments of the observed log-returns.

4.3 Algorithmic trading application.

4.3.1 Presentation of the strategy

The FNPH model may be helpful in practice for risk managers as well as traders. In this paper, we decided to apply it to the algorithmic trading framework. To do so, we simulate two portfolios. These portfolios are built on a trend-following strategy with the same signal. We add stop losses based on the forecasted volatility coming from the FNPH model from the E-GARCH model. The aim of this application is not to focus on portfolio returns, but to assess which model best reacts to turmoil periods. Here turmoil periods are defined by an excess of volatility. Each portfolio is defined as follows:

$$\begin{cases} \Pi_{t+1} = A_{t+1} + C_{t+1} \\ A_{t+1} = S_{t+1} Q_{t+1} \\ Q_{t+1} = Q_t + \mathcal{P}_{t+1} \\ C_{t+1} = C_t(1 + r_t) - \mathcal{P}_{t+1} S_{t+1} \end{cases}$$

where:

$$\begin{cases} \Pi_{t+1} \text{ is the value of the portfolio at time } t+1, \\ A_{t+1} \text{ is the value of the position in assets at time } t+1, \\ C_{t+1} \text{ is the cash reserve at time } t+1, \\ S_{t+1} \text{ is the value of the asset at time } t+1, \\ Q_{t+1} \text{ is the amount of asset owned at time } t+1, \\ \mathcal{P}_{t+1} \text{ is the traded amount of asset at time } t+1, \\ r_t \text{ is the risk-free rate at time } t. \end{cases}$$

In algorithmic trading, the trading decision at each time is based on a market indicator. Several forms of signal indicators exist, such as trend-following indicators (moving-average models, Bollinger bands), oscillation indicators (stochastic oscillator, Williams) or volume indicators (volume, on-balance volume). In this paper, the signal $z(t+1)$ is based on the difference between a short exponential moving average (EMA) and a long EMA. At each time $t+1$, we buy/sell a constant amount of index in function of the signal generated from the data until t. For sake of simplicity, we assume that the market is frictionless and that there is no tax.

To prevent portfolios from index crashes, stop-loss orders are added. Hence, if the investment signal, $z(t+1)$, is negative and if the forecasted volatility, $\hat{\sigma}_{t+1}$, is above a threshold $\mathcal{T}$ [15], then all the long positions in the risky asset are shorted.

We set:

$$\mathcal{P}_{t+1} = \begin{cases} 1 & \text{if } z(t+1) > 0 \text{ and if } \hat{\sigma}_{t+1} < \mathcal{T}, \\ -1 & \text{if } z(t+1) \leq 0 \text{ and if } \hat{\sigma}_{t+1} < \mathcal{T}, \\ Q_t & \text{if } z(t+1) \leq 0 \text{ and if } \hat{\sigma}_{t+1} \geq \mathcal{T}, \\ 0 & \text{else .} \end{cases}$$

Moreover, the strategy is forced to be long only, regarding the risky asset as well as the risk-free one. Therefore $A_t/\Pi_t \in [0,1]$ and $C_t/\Pi_t \in [0,1]$. Thus, no leverage is possible.

4.3.2 Data

To simulate portfolios, we use the French stock index CAC 40 as the risky asset, between January 11th, 2000 and August 5th, 2015. It represents 3979 trading days. The risk-free rate chosen is the

[15] We set $\mathcal{T}$ equals to the 98th percentile of the empirical distribution of volatility from January 3rd, 1990 until January 10th, 2000.

10-year French government bond. The E-GARCH model is re-estimated each day using a 400-day rolling window[16]. As for the forecast section, the aim of rolling-window techniques is to take into account transitory market changes. The FNPH model is estimated between January 3rd, 1990 and January 10th, 2000. As we previously said, we assess that g is structural and doesn't need to be re-estimated. The long EMA is estimated on the last 60 days whereas the short EMA is computed on the last 40 days.

4.3.3 Results

First we briefly compare the performance of the portfolios and then we focus on their reactivity to financial crisis.

Portfolios performance

To compare both the FNPH and the E-GARCH portfolios, we compute indicators used in asset management such as portfolio yearly returns, skewness, Sharpe ratio or maximum drawdown. For example, it is expected from efficient portfolios to have a profit and loss (PNL) with a positive asymmetry. The Sharpe ratio quantifies the excess of returns upon the risk exposed by the strategy. The risk is measured by the volatility. The maximum drawdown quantifies the largest peak-to-trough decline in the value of a portfolio. In other words, the maximum drawdown gives another measure of portfolio variability. Each portfolio starts on January 11th, 2000 with a million euro divided in 60% of cash and 40% of risky assets. We benchmark both portfolios against two other portfolios. One is remunerated along the trading period to the risk-free rate. The other one is only composed of shares of CAC 40 index. Table 6 presents an overview of the structure of the portfolios.

Portfolio	Trading Period	$\Pi(0)$	$\Pi(T)$
FNPH strategy	11/01/2000 to 5/08/2015	$1,000,000$	$2,237,068$
E-GARCH strategy	11/01/2000 to 5/08/2015	$1,000,000$	$1,831,241$
CAC 40	11/01/2000 to 5/08/2015	$1,000,000$	$920,100$
10-year OAT	11/01/2000 to 5/08/2015	$1,000,000$	$1,744,440$

Table 6: Portfolios structure.
Time 0 is January 11th, 2000 and time T is August 5th, 2015.

Figure 7 shows the portfolio values and Figure 8 the share of risky assets owned by each portfolio.

Table 7 provides strategy indicators. The Sharpe ratio benchmarked against the CAC 40 index is slightly higher for the FNPH portfolio than for the E-GARCH portfolio. As a comparison, the Sharpe ratio benchmarked against the risk-free asset indicates more clearly the better performance of the FNPH portfolio over the E-GARCH portfolio. This fact can be explained by the lower volatility of the FNPH portfolio compared to the E-GARCH portfolio. The FNPH portfolio is indeed farther from the risky asset than is the E-GARCH portfolio, so that the volatility of the returns in excess of the risky asset is higher in the case of the FNPH portfolio. So this higher volatility of the excess returns reduces the Sharpe ratio benchmarked against the CAC 40 index. The less risky feature of the FNPH portfolio also clearly appears when considering its maximum drawdown, which is much lower than the maximum drawdown of the E-GARCH portfolio. Indeed,

[16] In their article, Fang and Xu used a 200-day rolling window, but, as they wrote, they were focused on the first moment of log-returns [12]. As in this paper we are interested in forecasting the second moment of log-returns, we choose a longer rolling window

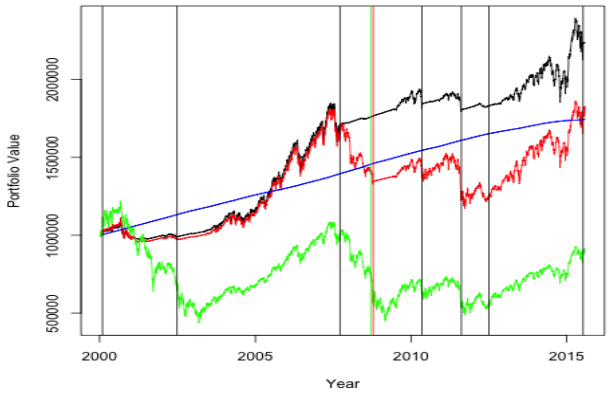


Figure 7: Simulated portfolios.
In black is FNPH portfolio. In red is E-GARCH portfolio. In blue 10-year OAT portfolio. In green is CAC 40 portfolio. The green vertical line indicates Lehman Brothers collapse, September 15th, 2008. The red vertical line indicates the activation of the stop-loss order for E-GARCH portfolio. The black vertical lines indicate the activation of stop-loss orders for FNPH-portfolio.

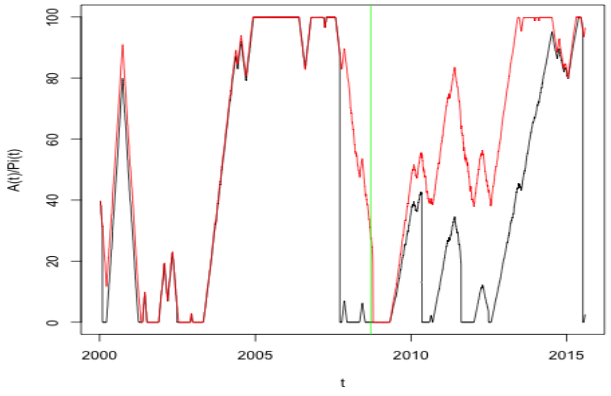


Figure 8: Simulated share (%) of risky assets owned by FNPH portfolio and E-GARCH portfolio.
In black is the share of risky asset of FNPH portfolio. In red is the share of risky asset of E-GARCH portfolio. The green vertical line indicates Lehman Brothers collapse, September 15th, 2008.

between the first peak and the lowest valley, the E-GARCH portfolio lost 44% of its value whereas the FNPH portfolio only lost 15% of its value. Furthermore, as we used a naive trading signal, the PNL of both portfolios is negatively asymmetric which means that losses occur more frequently than gains. Nevertheless, the skewness of FNPH portfolio is bigger than the one of E-GARCH portfolio. This is mainly due to the better reactivity of the FNPH model to turmoil periods.

	FNPH strategy	**E-GARCH strategy**
Skewness	−0.25	−0.42
Yearly rate of returns (%)	5.22	3.90
Annualized volatility (%)	7.50	10.01
Maximum drawdown (%)	14.97	43.66
Sharpe ratio benchmarked against CAC 40 (%)	32.82	30.13
Sharpe ratio benchmarked against 10-year OAT (%)	21.38	3.35
Number of stop-loss orders	7	1

Table 7: Strategy indicators.
The Sharpe ratio is built using the daily log-returns of the portfolio. The volatility is the annualized volatility of the daily log-returns of the portfolio built over the period. The annual number of trading days is set to 252.

Both portfolios exhibit a positive excess return. In addition the FNPH portfolio seems more reliable than the other portfolio, due to risk and performance considerations. However, it is worth noting that all the Sharpe ratios computed are quite low. This can be explained by at least three reasons.

- ▷ First, the buy/sell signal is computed using a naive indicator: EMA. This indicator is trend-following and so the signal can be improved using, for example, a mix with a mean-reverting indicator, or by including a forecasted volatility signal.
- ▷ Second, even if stop-losses prevent from turmoil periods and index crashes, they can generate fake positives, which lead to local losses. To correct this drawback, we can either use a more conservative quantile or sell out a smaller share of the position.
- ▷ Third, we can make the reactivity to market events better with an adaptive trading. For example, in case of a buy signal, the amount traded should be a decreasing function of the forecasted volatility. Indeed, when the forecasted volatility is high, then the standard deviation of log-returns is high, which means that the signal is uncertain.

Crisis periods and stop-loss orders

As we previously said, the forecasted volatility is used in both strategies as a signal for stop-loss orders. Then, when the forecasted volatility is above a certain threshold, all the positions of the portfolio in the risky asset are closed. Between 2001 and 2015, the two portfolios meet at least three crisis: the dot-com bubble crisis, the subprime crisis and the Lehman Brothers collapse and the European debt crisis.

Figure 8 shows the dynamical share of risky asset for both strategies. The share of risky asset in the FNPH portfolio is more sensitive to the market conditions than the share of risky asset in the E-GARCH portfolio. Furthermore, the FNPH portfolio generally looses less than the E-GARCH portfolio during crisis periods. For example, the E-GARCH portfolio reacted really late to the subprime crisis, as it can be seen in Figure 7: the E-GARCH portfolio lost 16.91% between June, 1st 2007 and September, 15th, 2008, whereas the FNPH portfolio only lost 5.67%. This

difference in term of reaction is generally explained by better forecast abilities of FNPH compared to E-GARCH.

In conclusion, the FNPH model is more robust to turmoil periods than the E-GARCH model.

5 Conclusion

In this paper we introduced a new method to model asset log-returns and volatility. Unlike all available time series model, such as GARCH-oriented processes, we do not require stationary log-returns. We recall that, without stationary constraints, GARCH processes are explosive. Besides, the increase of complexity of the latest GARCH improvements seems less limited than the associated better-fit benefits. The model we proposed, is fully non-parametric and it is based on innovative mathematical tools, such as wavelet denoising or variational theory. Furthermore, the model reproduces the sytilized facts of log-returns and it has better forecast abilities than GARCH-oriented processes.

Some improvements remain to be done:

- ▷ First, the vector of parameters (μ, δ) is chosen according to a quite arbitrary criterion. We proposed a naive method to make this choice easy, but alternative methods could be tested.
- ▷ Second, the model has only been computed for a time lag, h, equal to 1. Thus, a momentum effect may be forgotten. To overcome this limitation, algorithms close to nearest neighbours or solving travelling salesman problem could be used.
- ▷ Third, this model could be extended to the multivariate case.

A Justification of equation (3)

We consider the estimation problem of g, from the model:

$$y(t) - x_i(t) = g(\mathcal{Y}(t))\varepsilon_t.$$

Let δ_ε be the probability density function of a unit Gaussian random variable:

$$\delta_\varepsilon : e \in \mathbb{R} \mapsto \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{e^2}{2}\right).$$

Since g is assumed to be positive, we can apply the standard relation [2]:

$$\delta_\epsilon\left(\frac{y(t) - x_i(t)}{g(\mathcal{Y}(t))}\right)\frac{1}{g(\mathcal{Y}(t))} = \delta_{y(t)-x_i(t)|g(\mathcal{Y}(t))}(y(t) - x_i(t)|g(\mathcal{Y}(t))). \tag{4}$$

The maximum-likelihood problem consists in maximizing the right-hand side of equation (4). It is therefore equivalent to minimizing the opposite of the logarithm of the left-hand side of the same equation, that is, excluding constant terms, for each time t:

$$\log(g_{i+1}(\mathcal{Y}(t))) + \frac{1}{2}\left(\frac{y(t) - x_i(t)}{g_{i+1}(\mathcal{Y}(t))}\right)^2.$$

Summing that function over all the observations by the path θ leads to the following continuous form of the minimization problem:

$$\int_0^T \left[\log(\mathcal{G}_{i+1}(t)) + \frac{1}{2} \left(\frac{y(\theta(t)) - x_i(\theta(t))}{\mathcal{G}_{i+1}(t)} \right)^2 \right] dt,$$

where $\mathcal{G}_{i+1} = g_{i+1} \circ \mathcal{Y} \circ \theta$. Moreover, we impose a condition of smoothness for g, as an additional objective of minimizing its quadratic variations over the path θ. Therefore, we now aim to minimize, for each time t:

$$\int_0^T \mathcal{L} \left(t, \mathcal{G}_{i+1}(t), \frac{d}{dt}\mathcal{G}_{i+1}(t) \right) dt,$$

where

$$\mathcal{L} \left(t, \mathcal{G}_{i+1}(t), \frac{d}{dt}\mathcal{G}_{i+1}(t) \right) = \mu \left[\log(\mathcal{G}_{i+1}(t)) + \frac{1}{2} \left(\frac{y(\theta(t)) - x_i(\theta(t))}{\mathcal{G}_{i+1}(t)} \right)^2 \right] + \frac{1}{2} \left(\frac{d}{dt}\mathcal{G}_{i+1}(t) \right)^2,$$

where $\mu > 0$ is a given parameter. $\mathcal{G}_{i+1}$ is therefore the solution of the corresponding Euler-Lagrange equation:

$$\begin{aligned} 0 &= \frac{\partial}{\partial \mathcal{G}_{i+1}} \mathcal{L} \left(t, \mathcal{G}_{i+1}(t), \frac{d}{dt}\mathcal{G}_{i+1}(t) \right) - \frac{d}{dt} \frac{\partial}{\partial \frac{d}{dt}\mathcal{G}_{i+1}} \mathcal{L} \left(t, \mathcal{G}_{i+1}(t), \frac{d}{dt}\mathcal{G}_{i+1}(t) \right) \\ &= \mu \left[\frac{1}{\mathcal{G}_{i+1}(t)} - \frac{(y(\theta(t)) - x_i(\theta(t)))^2}{\mathcal{G}_{i+1}(t)^3} \right] - \frac{d^2}{dt^2}\mathcal{G}_{i+1}(t). \end{aligned}$$

It justifies equation (3).

B Results

	FNPH	GARCH(1,1)	E-GARCH(1,1)	GARCH-GJR(1,1)
EUR/GBP	10109	9551	9643	9572
EUR/USD	9845	9231	9304	9235
CAC 40	8520	7859	7864	7870

Table 8: Log-likelihoods of the estimated processes.
For exchange-rate data, log-likelihoods are computed on the first 1500 observations. For CAC 40 data, log-likelihoods are computed on the first 2500 observations.

Dataset		**Model**	MAD	RMSE
EUR/GBP	Volatility	GARCH(1,1)	$2.13e-4$	$3.00e-4$
		E-GARCH(1,1)	$2.10e-4$	$2.90e-4$
		GARCH-GJR(1,1)	$2.10e-4$	$3.02e-4$
		FNPH	$1.93e-4$	$2.81e-4$
	Log-returns	GARCH(1,1)	$2.57e-4$	$3.74e-4$
		E-GARCH(1,1)	$2.58e-4$	$3.74e-4$
		GARCH-GJR(1,1)	$2.56e-4$	$3.56e-4$
		FNPH	$2.52e-4$	$3.65e-4$
EUR/USD	Volatility	GARCH(1,1)	$2.66e-4$	$3.84e-4$
		E-GARCH(1,1)	$2.69e-4$	$3.93e-4$
		GARCH-GJR(1,1)	$2.70e-4$	$3.93e-4$
		FNPH	$2.33e-4$	$3.42e-4$
	Log-returns	GARCH(1,1)	$3.18e-4$	$4.80e-4$
		E-GARCH(1,1)	$3.21e-4$	$4.82e-4$
		GARCH-GJR(1,1)	$3.15e-4$	$4.76e-4$
		FNPH	$3.10e-4$	$4.69e-4$
CAC 40	Volatility	GARCH(1,1)	$7.04e-3$	$1.02e-2$
		E-GARCH(1,1)	$6.94e-3$	$1.01e-2$
		GARCH-GJR(1,1)	$6.69e-3$	$1.00e-2$
		FNPH	$6.76e-3$	$1.04e-2$
	Log-returns	GARCH(1,1)	$9.12e-3$	$1.28e-2$
		E-GARCH(1,1)	$9.14e-3$	$1.29e-2$
		GARCH-GJR(1,1)	$9.13e-3$	$1.29e-2$
		FNPH	$8.89e-3$	$1.25e-2$

Table 9: Forecast errors of other datasets for the four models.
For exchange-rate data, forecasts are done on the last 291 observations. For CAC 40 data, forecasts are done on the last 3979 observations.

Dataset	GARCH(1,1)	E-GARCH(1,1)	GARCH-GJR(1,1)
EUR/GBP	37.68	44.77	50.74
EUR/USD	29.17	30.92	24.73
CAC 40	0.74	0.42	0.35

Table 10: Weighted likelihood ratio.
The ratio is computed between the FNPH model and the three GARCH-oriented models. For exchange-rate data, the statistic is computed on the 291 density forecasts. For CAC 40 data, the statistic is computed on the 3979 density forecasts

References

[1] Amisano, G. and Giacomini, R. (2007), *Comparing density forecasts via weighted likelihood ratio tests*, Journal of Business and Economic Statistics, 25, 2: 177-ĂŞ190

[2] Aubert, G. and Aujol, J.-F. (2008), *A variational approach to removing multiplicative noise*, SIAM Journal on Applied Mathematics, 68, 4: 925-946

[3] Bollerslev, T (1986), *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, 31, 3: 307-327

[4] Buhlmann, P. and McNeil, A. J. (2002), *An algorithm for nonparametric GARCH modelling*, Computational Statistics & Data Analysis, 40, 4: 665-683

[5] Cont, R. (2001), *Empirical properties of asset returns: stylized facts and statistical issues*, Quantitative Finance, 1: 223-226

[6] Dahlhaus, R. and Rao, S. S. (1986), *Statistical inference for time varying ARCH processes*, The Annals of Statistics, 34, 3: 1074-1114

[7] Davis, R.A. and Mikosch, T. (2008), *Extreme value theory for GARCH processes*, in Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (editors) *Handbook of Financial Time Series*, Springer, New-York

[8] Donoho, D. and Johnstone, I. (1994), *Ideal spatial adaptation via wavelet shrinkage*, Biometrika, 81, 3: 425-455

[9] Donoho, D. and Johnstone, I. (1995), *Adapting to unknown smoothness via wavelet shrinkage*, Journal of the American statistical association, 90, 432: 1200-1244

[10] Engle, R. F. (1982), *Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom Inflation*, Econometrica, 50: 987-1008

[11] Feng, Y. (2004), *Simultaneously modelling conditional heteroskedasticity and scale change*, Econometric Theory, 20, 3: 563-596

[12] Fang, Y. and Xu, D. (2003), *The predictability of asset returns: an approach combining technical analysis and time series forecasts*, International Journal of Forecasting, 19, 3: 369-385

[13] Garcin, M. (2015), *Empirical wavelet coefficients and denoising of chaotic data in the phase space*, in Skiadas, C.H. and Skiadas, C. (editors), *Handbook of applications of chaos theory*, to appear in CRC/Taylor & Francis

[14] Garcin, M. and Guégan, D. (2014), *Probability density of the empirical wavelet coefficients of a noisy chaos*, Physica D: Nonlinear Phenomena, 276: 28-47

[15] Garcin, M. and Guégan, D. (2014), *Optimal wavelet shrinkage of a noisy dynamical system with non-linear noise impact*, working paper

[16] Glosten, L. R., Jagannathan, R. and Runkle, D. (1993), *On the relationship between the expected value and the volatility of the nominal excess return on stocks*, Journal of Finance, 48, 5: 1779-1801

[17] Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized additive models*, Chapman & Hall/CRC, London, UK

[18] Hagan, P., Kumar, D., Lesniewski, A. and Woodward, D. (2002), *Managing the smile risk*, Wilmott Magazine

[19] Heston, S. L. (1993), *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, The Review of Financial Studies, 6, 2: 327-343

[20] Kanaya, S. and Kristensen, D. (2010), *Estimation of stochastic volatility models by non-parametric filtering*, Working paper

[21] Mallat, S. (2000), *Une exploration des signaux en ondelettes*, Ellipses, Éditions de l'École Polytechnique, Paris, France

[22] Mandelbrot, B. (1963), *The variation of certain speculative prices*, The Journal of Business, 36: 392-417

[23] Mikhailov, S. and Nogel, U. (2003), *Heston's stochastic volatility model. Implementation, calibration and some extensions*, Wilmott Magazine: 74-94

[24] Mishra, S., Su, L. and Ullah, A. (2010), *Semiparametric estimator of time series conditional variance*, Journal of Business and Economic Statistics, 28, 2: 256-274

[25] Nelson, D. B. (1991), *Conditional heteroskedasticity in asset returns: a new approach*, Econometrica, 59, 2: 347-370

[26] Poon, S. and Granger, C W. (2005), *Practical issues in forecasting volatility*, Financial Analysts Journal, 61, 1: 45-56

[27] Stein, C. (1981), *Estimation of the mean of a multivariate normal distribution*, The Annals of Statistics, 9, 6: 1135-1151

[28] Tassi, Ph. (1989), *Méthodes Statistiques*, Economica, 2nd edition, Paris, France

[29] Wang, L., Feng, C., Song, Q. and Yang, L. (2012), *Efficient semiparametric GARCH modelling of financial volatility*, Statistica Sinica, 22, 1: 249-270

[30] Winker, P. and Maringer, D. (2009), *The convergence of estimators based on heuristics: theory and application to a GARCH model*, Computational Statistics, 24, 3: 533-550