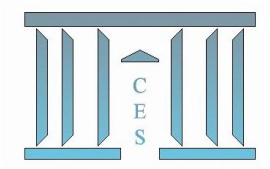
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers



Non-parametric news impact curve: a variational approach

Matthieu GARCIN, Clément GOULET

2015.86R

Version révisée



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

Non-parametric news impact curve: a variational approach

Matthieu Garcin*, Clément Goulet†‡

July 11, 2016

Abstract

In this paper, we propose an innovative methodology for modelling the news impact curve. The news impact curve provides a non-linear relation between past returns and current volatility and thus enables to forecast volatility. Our news impact curve is the solution of a dynamic optimization problem based on variational calculus. Consequently, it is a non-parametric and smooth curve. To our knowledge, this is the first time that such a method is used for volatility modelling. Applications on simulated heteroskedastic processes as well as on financial data show a better accuracy in estimation and forecast for this approach than for standard parametric (symmetric or asymmetric ARCH) or non-parametric (Kernel-ARCH) econometric techniques.

1 Introduction

For the last three decades, various models have been proposed for forecasting volatility. Volatility has many applications in finance and its forecast is indeed useful in risk measurement, portfolio management, trading strategies or options pricing. One way of forecasting volatility is to take advantage of its interaction with returns. The news impact curve defines the relation between past returns and current volatility [13]. This relation has been well described and presents two notable characteristics. First, it is a function that decreases for negative past returns and then increases for positive past returns. It thus reaches its minimum around 0: weak returns, in absolute value, induce a weak uncertainty, whereas bigger returns, in absolute value, lead to higher uncertainty. Second, it is an asymmetric function. One explanation is that "bad news" (high negative returns) create more volatility than "good news" (high positive returns). Hence, volatility models have to take account of these two features. For example, ARCH models impose a symmetric and hyperbolic relation between past returns and current conditional volatility [12].

Parameters can be added to the news impact curve to take into account returns asymmetry. GJR-GARCH [19] and E-GARCH [30] are such parametric asymmetric models. They are extensions of the GARCH model, which has been introduced by Bollerslev [5] and which depicts the self-dependence of the volatility across time. In other words, GARCH models consider volatility as

*Natixis Asset Management - LabEx ReFi

†University Paris 1 - LabEx ReFi

‡We thanks for their valuable comments: Christophe Boucher, Éric Gautier, Dominique Guégan, Florian Ielpo, Philippe de Peretti and Todd Prono. This acknowledgement does not make them accountable for any remaining mistake. This work was achieved through the Laboratory of Excellence on Financial Regulation (Labex ReFi) supported by PRES heSam under the reference ANR10LABX0095. " It benefited from a French government support managed by the National Research Agency (ANR) within the project Investissements d'Avenir Paris Nouveaux Mondes (invesments for the future ParisNew Worlds) under the reference ANR11IDEX000602.

a weighted sum of a news impact curve and of past volatility realizations. All these parametric hyperbolic news impact curves have some limits. First, adding parameters for better matching statistical properties of volatility increases the estimation complexity. As a consequence, their estimation requires a large number of observations to converge. Then, since self-dependence smooths the volatility, GARCH-oriented models are quite slow to react to extreme market moves.

Semi-parametric and non-parametric approaches overpass the constraint of a hyperbolic relation between volatility and past returns. Pagan and Schwert (1990) developed a smooth function, based on a Nadaraya-Watson estimator, to model conditional variance [32]. Gouriéroux and Montfort (1992) are the first to build a semi-parametric equation (QTARCH) to model conditional mean and conditional variance [20]. Their approach mixes Markovian theory and parametric step functions. Härdle and Tsybakov (1997) extended their work to a larger class of step functions [21]. Linton and Mammen (2005) built a semi-parametric estimate of the news impact curve using local linear estimators [25]. This last model differs from others by the inclusion of the volatility persistence. Other approaches have been developed to include the long memory property of volatility and non-parametric news impact curves [22, 23, 24].

Most of the mentioned methods are based on non-parametric regressions and estimators rely on the minimization of some quadratic error. In the present paper, instead of estimating the news impact curve by minimizing a quadratic error, we maximize a sum of likelihoods over all the observations. It allows a higher consistence between the empirical distribution of the innovations and what is assumed by the model. Moreover, in order to avoid overfitting, we impose a smoothness constraint on the news impact curve. Therefore, our problem consists in minimizing a sum over all the observations of a functional of the news impact curve and of its increments. The variational calculus theory provides a natural solution to this problem. The literature in economics and finance mentions various applications of this theory, for example for optimal income allocation [36], hedge fund asset allocation [8] or studies of options price sensitivities with stochastic variational calculus (Malliavin calculus) [27].

The paper is organized as follows. In section 2, we present our model and a short algorithm for its estimation. In section 3, we apply it on simulated data and on three financial sets of log-returns. Our model is compared to four standard parametric and non-parametric models. To compare out-of-sample forecast properties, we use the Diebold-Mariano-West procedure [9, 37]. On all the datasets tested, our model presents globally better results than ARCH, NP-ARCH, GARCH and GJR-GARCH, in and out of sample.

2 Model and estimation

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

$$\left\{ \begin{array}{rcl} y_t & = & x_t + \varepsilon_t \\ \varepsilon_t & = & \sqrt{h_t} z_t \\ \sqrt{h_t} & = & g(\varepsilon_{t-1}, ..., \varepsilon_{t-l}), \end{array} \right. \tag{1}$$

in which y is the observed price return of an asset, x is an unknown deterministic function corresponding to the return of the fundamental asset value, $l \geq 1$ is an integer indicating the number of lags in the information and ε_t is the noisy part of the observed price at time t. More precisely, for every t, the innovation z_t is a unit Gaussian random variable which is independent of z_s, for every $s \neq t$, and of x_u, for every u. Moreover, g is an unknown deterministic and positive function, such that $g(\mathcal{E}_t) = \sqrt{h_t}$, where $\mathcal{E}_t$ is the lagged information of past residuals, $(\varepsilon_{t-1}, ..., \varepsilon_{t-l})$. We can easily compute the first unconditional moment and the conditional variance of the model defined

by equation (1):

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

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

In this model, the conditional standard deviation of returns is a function of past innovations. This model is a very general form of an ARCH-type model with time-varying trend. We could set a parametric form to g, to recover a standard ARCH model but we prefer to define a non-parametric framework both for the trend, x, and for the news impact curve, g.

First, the trend x_t is estimated by a wavelet denoising approach. Wavelets are used to decompose a signal, here y, in different frequencies. In particular, it can be written as the sum of a gross structure and details, which are often very erratic. An important part of these details corresponds to a noise that we want to eliminate. Standard wavelet denoising rules enable to minimize the quadratic error between the pure and unknown trend x and its estimate. These rules stipulate that a wavelet coefficient[1] should be above a certain threshold if it is relevant: it must be kept as such. On the contrary, small coefficients are supposed to be mostly due to a noise: they can be shrunk to zero.

Second, we use a variational approach for estimating the news impact curve. Such a method is not usual in econometrics. However, its use naturally arises as we explain in the following lines. For clarity purposes, we only focus on the case of a unique delay in g, just like in an ARCH(1) model. We now assume that x has first been estimated.[2] Then g is such that :

$$g(\varepsilon_{t-1})z_t = y_t - x_t. \tag{2}$$

Given ε_{t-1}, we can estimate $g(\mathcal{E}_t) = g(\varepsilon_{t-1})$ by maximizing a likelihood. The set of the sorted lagged observed innovations $\mathcal{E}_{\theta(0)} \leq \mathcal{E}_{\theta(1)} \leq ... \leq \mathcal{E}_{\theta(T)}$, where θ is an ordering function, form a discretization of an interval of $\mathbb{R}$. Therefore, we estimate g by maximizing the log-likelihood corresponding to equation (2) at each time, that is we maximize the log-likelihood over the whole time interval:

$$\sum_{t=0}^{T} \tilde{\mathcal{L}}(t, \mathcal{G}(t)), \tag{3}$$

where we set $\mathcal{G}(t) = g(\mathcal{E}_{\theta(t)}) = g(\varepsilon_{\theta(t)-1})$ for each t and where $\tilde{\mathcal{L}}$ is the log-likelihood. The exact expression of $\tilde{\mathcal{L}}$, for a Gaussian noise, is given by proposition 1, whose proof is given in Appendix A.4.

Proposition 1. *The log-likelihood of $y_{\theta(t)} - x_{\theta(t)}$ given $\mathcal{G}(t)$, for a Gaussian noise, is:*

$$\tilde{\mathcal{L}}(t, \mathcal{G}(t)) = C + \log(\mathcal{G}(t)) + \frac{1}{2}\left(\frac{y_{\theta(t)} - x_{\theta(t)}}{\mathcal{G}(t)}\right)^2,$$

where C is a constant term.

In a continuous setting,[3] equation (3) would become:

$$\int_0^T \tilde{\mathcal{L}}(t, \mathcal{G}(t))dt. \tag{4}$$

[1] That is the part of y decomposed on a particular frequency at a particular date.

[2] Our method is iterative: we alternate the estimation of x given g and of g given x. We initiate the iteration by the estimation of x for a basic constant volatility term g.

[3] Our model is written in equation (1) in discrete time, with the time t in a subset of integers, like in most of the literature about ARCH models. However, some papers deal with the continuous limit of ARCH models [29] or

Up to now, maximizing such an integral is similar to maximizing the integrand and it does not need any unusual technique. However, by doing so, the estimated value of $\mathcal{G}(t)$ will be disconnected from the estimated value of $\mathcal{G}(s)$, for any $s \neq t$. This can provide an erratic news impact curve. Moreover, such a method would lead to low in-sample errors and high out-of-sample errors. Such an occurrence of overfitting is often associated to non-parametric models. We get past it by adding a smoothing term in equation (4), which aims at minimizing the quadratic variation of $\mathcal{G}$. We thus define a new functional form

$$\mathcal{L}\left(t, \mathcal{G}(t), \frac{d}{dt}\mathcal{G}(t)\right) = \mu\tilde{\mathcal{L}}(t, \mathcal{G}(t)) + \frac{1}{2}\left(\frac{d}{dt}\mathcal{G}(t)\right)^2,$$

so that we now look for an extremum of the equation

$$\int_0^T \mathcal{L}\left(t, \mathcal{G}(t), \frac{d}{dt}\mathcal{G}(t)\right)dt. \tag{5}$$

The optimization problem defined in equation (5) cannot be solved as successive independent optimization problems anymore. We thus need the help of the variational calculus. In particular, we use the Euler-Lagrange equation, so as to transform the optimization of an integral in several interrelated optimizations indexed by t.

Proposition 2. *Let $\mathcal{G}$ be two times differentiable and $\mathcal{L}$ be one time differentiable and defined by equation (5). $\hat{\mathcal{G}}$ allows to reach an extremum of $\mathcal{L}$ if and only if $\hat{\mathcal{G}}$ is solution of the Euler-Lagrange equation:*

$$\forall t \in (0,T),\ 0 = \frac{\partial}{\partial\mathcal{G}}\mathcal{L}\left(t, \mathcal{G}(t), \frac{d}{dt}\mathcal{G}(t)\right) - \frac{d}{dt}\frac{\partial}{\partial\frac{d}{dt}\mathcal{G}}\mathcal{L}\left(t, \mathcal{G}(t), \frac{d}{dt}\mathcal{G}(t)\right).$$

This proposition is a standard result of the variational theory. A proof can be found in [18]. This equation leads to a concise optimization problem, which finally enables to get an estimate of $\mathcal{G}$ and g.

From a practical point of view, we iterate the estimation of x and the estimation of g, since the estimate of g is used in x and *vice versa*. The algorithm can be summarized as the following pseudo-code:[4]

---

even of GARCH models [3, 7]. In such a continuous framework, our model would be the limit, when $\tau \to 0$, of

$$\left\{\begin{array}{rcl} y_{t\tau} & = & x_{t\tau} + \varepsilon_{t\tau} \\ \varepsilon_{t\tau} & = & \sqrt{h_{t\tau}}z_{t\tau} \\ \sqrt{h_{t\tau}} & = & g(\varepsilon_{(t-1)\tau}, ..., \varepsilon_{(t-l)\tau}), \end{array}\right.$$

for t still an integer. We thus indifferently write a discrete sum like in equation (3), which is consistent with the classical ARCH models, or a continuous integral like in equation (4), which is consistent with the variational framework. In particular, when we write a derivative $d\mathcal{G}(t)/dt$ in this continuous setting, like in equation (5), it must be interpreted, in the basic discrete framework, as $(\mathcal{G}(t) - \mathcal{G}(t-\tau))/\tau$ with $\tau = 1$.

[4] These lines of pseudo-code only present the main architecture of the algorithm. They refer to functions with explicit name, which also exist in many programming language but under another name. Specifically, $GetWaveletCoefficients$ creates a vector of wavelet coefficients, $Median$ calculates a median, $GetFilteredCoefficients$ applies a threshold filter to wavelet coefficients, $GetWaveletReconstruction$ computes an inverse wavelet transform and $GetOrderingIndexes$ provides the permutation allowing to sort in ascending order the coordinates of a vector.

```
1  WaveletCoefficients = GetWaveletCoefficients(Y)
2  G = Median(WaveletCoefficients)/0.6745
3  for(i = 1; i ≤ NumberIteration1; i++)
4        FilteredWaveletCoefficients =
               GetFilteredCoefficients(WaveletCoefficients, NoiseAmplitude = G)
5        X = GetWaveletReconstruction(FilteredWaveletCoefficients)
6        Theta = GetOrderingIndexes(Y − X)
7        for(n = 1; n ≤ NumberIteration2; n++)
8              G = Median(WaveletCoefficients)/0.6745
9              for(t = 1; t <= T; t++)
10                   G(t) = G(t) + delta ∗ (G(t + 1) − 2 ∗ G(t) + G(t − 1)
                           −mu ∗ (G(t)^2 − (Y(Theta(t)) − X(Theta(t)))^2)/G(t)^3)
```

The algorithm for estimating g is more detailed in Appendix A, as well as some possible refinements exposed in Appendix A.3. We call our new model the wavelet-variational ARCH, later WV-ARCH in this paper.

3 Applications

3.1 Simulated data

In this section, we compare estimation and out-of-sample performances of the WV-ARCH model with ARCH and NP-ARCH models, for a simulated process $\{s_t, t \in \mathbb{N}\}$. We get the simulated process from Linton and Mammen (2005), which compared the estimation accuracy of non-parametric news impact curves with a simulated process [25]. Following this, we impose an asymmetric volatility process:

$$\left\{ \begin{array}{rcl} s_t & = & \sqrt{h_t} z_t \\ h_t & = & \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} + \theta 1_{\varepsilon_{t-1}<0} \varepsilon_{t-1}^2, \end{array} \right. \tag{6}$$

with $\{z_t\}$ a set of independent random unit Gaussian variables and with the same values for the parameters than in the mentioned paper: $\omega = 0.2$, $\alpha = 0.06$, $\beta = 0.9$ and $\theta = 0.03$.

We estimate on the first 1000 observations a WV-ARCH(1), an ARCH(1) and a non-parametric ARCH(1) model. The WV-ARCH model is estimated by the algorithm exposed in the previous section but without the wavelet part: no drift is assumed, neither for WV-ARCH nor for the other models. The ARCH model is estimated by a maximum-likelihood method and the non-parametric ARCH by the Pagan-Schwert procedure [32]. The accuracy of the estimation of the three models is gauged by the metrics gathered in Table 1.

	Mean	Variance	Skewness	Excess kurtosis	Log-lik.	K.-S. p-value
ARCH	$3.47e-2$ *(23.8%)*	0.999 *(85.7%)*	0.050 *(51.7%)*	2.10 *(<0.1%)*	-1418.5	3.6%
NP-ARCH	$1.51e-2$ *(55.7%)*	0.998 *(83.9%)*	-0.125 *(10.4%)*	1.44 *(<0.1%)*	-1417.4	16.8%
WV-ARCH	$3.46e-2$ *(23.9%)*	0.993 *(76.0%)*	-0.016 *(83.0%)*	-0.32 *(2.0%)*	-1415.9	47.8%

Table 1: Four first moments of the innovations, as well as log-likelihood of the estimated model and p-value of a normality test for the innovations. Models are estimated on the first 1000 observations of the simulated process $\{s_t, t \in \mathbb{N}\}$. The normality test is a Kolmogorov-Smirnov test. Under the four first moments is indicated the p-value of the moment for the null hypothesis of a standard Gaussian distribution. The tests are respectively: T-test, F-test, D'Agostino test, Anscombe-Glynn test.

Since the random variables z_t in the simulated process are Gaussian, then the innovations are expected to be close to unit Gaussian random variables. In Table 1, we observe that the WV-ARCH innovation distribution better fits a $\mathcal{N}(0,1)$-distribution than do ARCH and NP-ARCH. The innovations of each of the three models have a mean close to zero, and the innovations of WV-ARCH and ARCH have a variance close to 1. Moreover, while the innovations of both ARCH and NP-ARCH are leptokurtic and quite asymmetric, WV-ARCH's skewness and kurtosis are close to zero. Globally, the higher Gaussianity of WV-ARCH's innovations is highlighted by the higher likelihood and by the larger p-value for a normality test.

We also make out-of-sample forecasts. WV-ARCH, ARCH and NP-ARCH are re-estimated at each time step using a rolling window of length $H = 500$. The volatility is not directly observable. We thus have to choose a proxy measure of the volatility to estimate the out-of-sample accuracy of the model. This measure is a benchmark for our volatility forecasts. Two non-parametric measures are commonly used in the volatility literature: the realized volatility and the absolute-return volatility [38]. As absolute returns are a noisy measure of volatility, we choose a kernel-estimator proxy of realized volatility [4].

To compare by pairs the forecast accuracy between the three models, we choose a criterion reflecting the accuracy, that is the ability of the predicted volatility, $\hat{h}_t$, to depict the proxied realized volatility, σ_t. It is done with a so-called loss function, which satisfies two conditions: to be robust both to the noise induced by the volatility proxy and to the conditional distribution of log-returns (especially their first two moments). Patton and Sheppard [33] demonstrated that the QLIKE loss function satisfies these two conditions. The QLIKE loss function is defined by:

$$L(\sigma_t^2, \hat{h}_t) = -1 - log\left(\frac{\sigma_t^2}{\hat{h}_t}\right) + \frac{\sigma_t^2}{\hat{h}_t}.$$

The higher L, the less accurate the model.

We then compare the volatility forecasts of two models using a Diebold-Mariano-West test (DMW)[9, 37] with a QLIKE loss function. WV-ARCH, ARCH are NP-ARCH are estimated on a rolling window of size $H = 500$. We thus make a first forecast in time $H + 1$. We iterate by estimating the model on the data at times t to $t + H - 1$ and make a forecast for time $t + H$, t ranging from 1 to $T = 500$. The DMW test statistic for the T forecasts of each of two models, DMW_T, is computed by taking the difference of the loss functions of the two models:

$$d_T = \frac{1}{T}\sum_{t=H+1}^{H+T}\Big[L(\sigma_t^2, \hat{h}_{t,1}) - L(\sigma_t^2, \hat{h}_{t,2})\Big],$$

where $\hat{h}_{t,i}$ is the forecast provided by model i. By definition:

$$DMW_T = \frac{\sqrt{T}d_T}{\sqrt{\hat{avar}(\sqrt{T}d_T)}},$$

where $\hat{avar}$ is the Newey-West long-run variance estimator of the re-scaled sample mean $\sqrt{T}d_T$ [31].

Table 2 presents the results of DMW tests for the two models compared to WV-ARCH. Among all the tested models, WV-ARCH has the smallest QLIKE loss function. Its forecast ability is significantly higher since, in addition, the test statistic of DMW test permits to reject, respectively at 99% and 90% confidence level, the equality of losses with ARCH and NP-ARCH.

Model	QLIKE	DMW_T *vs* WV-ARCH
ARCH	$9.45e-2$	4.00***
NP-ARCH	$7.27e-2$	1.71*
WV-ARCH	$6.72e-2$	-

Table 2: QLIKE losses and DMW *vs* WV-ARCH statistics. *, ** and *** signify rejecting the null hypothesis of equal losses for respectively 90%, 95% and 99% confidence levels.

In conclusion, WV-ARCH has better estimation results and shows a better out-of-sample forecast accuracy on simulated data than the ARCH and NP-ARCH models.

3.2 Financial data

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

In the applicative part of this paper, we only provide g for $l = 1$. The main challenge of the WV-ARCH model is to provide a relevant estimate of g. As the theoretical form g for an observable system is unknown, we cannot compare graphically the estimate $\hat{g}$ to g. To overcome this issue, we compare each estimated news impact curve $\hat{g}$ to an estimate σ_t of the realized volatility[5].

Figure 1 shows an example of the estimation of g for S&P 500 log-returns between the 3rd of January 2000 and the 24th of August 2007. For the three log-return series used, we obtain a relationship between the instantaneous innovations and the one-time-ahead volatility which captures both returns asymmetry and volatility clustering. The clustering effect occurs because high returns in absolute value are followed by high volatility. The asymmetry is striking in Figure 1, in which the distance between our non-parametric news impact curve and the asymptotic line of the ARCH hyperbolic news impact curve is much bigger on the left than on the right. For the particular case of stock returns, the returns asymmetry can be interpreted as the leverage effect.

To validate the model we designed, we compare it to the classical time series framework. We treat four well-known models, whose equations are given in Table 3: the ARCH model [12], the non-parametric ARCH model based on Nadaraya-Watson kernel estimator [32], GARCH(1,1) [5] and the GJR-GARCH(1,1) [19]. To improve the accuracy and the fairness of the comparisons, we replaced the constant drift term in all the time series model by a moving drift term, x_t, obtained by wavelet shrinkage.

[5] An estimate of the realized volatility for S&P 500, FTSE 100 and DAX is available on *Oxford-Man Institute*

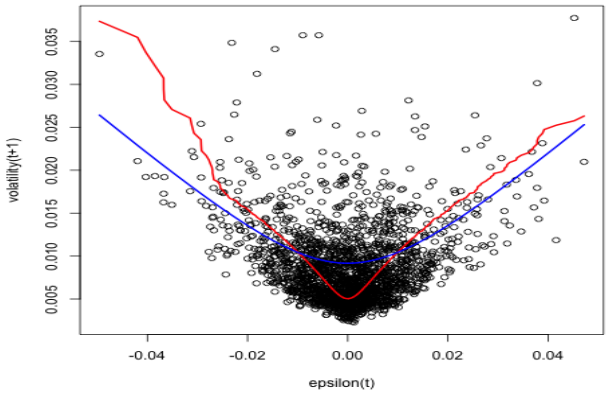


Figure 1: Estimates of g for S&P 500 log-returns.
The bullet points are the realized volatility, the red line is the estimate of g after one iteration in the first loop indexed by i of the pseudocode presented in section 2. The blue line is the news impact curve estimated for an ARCH(1) model.

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

Table 3: Time series models.
Each model is defined by three equations: the first two are shared by all the models ($y_t = x_t + \varepsilon_t$ and $\varepsilon_t = \sqrt{h_t}z_t$), the third one is specific and defines the volatility, as written in the table. All the innovation processes $(z_t)_t$ are assumed to be standard Gaussian variables. In ARCH-NP, m is a kernel function.

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

3.2.1 In-sample estimation

To find the model that better reproduces the stylised facts of log-returns, we compare their residuals. The purpose of this study is to find whether our model better matches the statistical properties of financial assets log-returns. Various papers have studied these statistical properties [28, 6, 34]. We focus on five stylised facts: conditional heteroskedasticity, volatility clustering, leverage effects, fat tails and asymmetric distributions. A model reflecting those stylised facts should have residuals consistent with the distribution specified in the model. It is thus expected from the innovations of all these Gaussian models to have moments close to those of a unit Gaussian. In particular, we focus on the four first moments. We also expect that the innovations pass a Gaussian test. And finally, a significantly better model will have a higher likelihood. We gather those statistics for the five models on S&P 500, FTSE 100 and DAX in Table 3.2.1.

The estimates of the WV-ARCH model are obtained after only one iteration of our first loop, indexed by i, in the pseudocode provided in Section 2.[6] By doing so, all the tested models have the same estimated trend x_t, which does not depend on our non-parametric news impact curve g.

The innovations of the WV-ARCH model seem globally close to a standard Gaussian distribution. More precisely, among the three ARCH models, the innovations of the WV-ARCH model are the closest to a Gaussian variable, regarding the mean, the skewness, the kurtosis and the log-likelihood. Except for the FTSE 100, The p-value of the Kolmogorov-Smirnov test on the innovations is always higher for the WV-ARCH than for ARCH and NP-ARCH models. In particular, the normality is not rejected for the WV-ARCH, whereas it is always rejected for ARCH and NP-ARCH. The difference with innovations of GARCH-oriented model mitigates the superiority of the WV-ARCH model. This is due to the introduction of the persistence of the volatility, β. Across all the datasets, only the mean, the skewness and the log-likelihood unanimously state a higher closeness of the $\mathcal{N}(0,1)$ distribution with WV-ARCH innovations than with GARCH and GARCH-GJR innovations. This result highlights the ability of WV-ARCH to model log-returns asymmetry.

3.2.2 Out-of-sample forecasts

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

Table 5 presents the results of DMW tests for the five models.

Forecasts are done during 400 trading days, between the 27th of August 2007 and the 6th of March 2009, so that it includes periods of high and low volatility. Models are re-estimated at each time step using a rolling-window procedure. The size of the rolling window is set to 1000 trading days. In Table 5 we observe that, for each series of log-returns, the null hypothesis of equality of QLIKE loss function is always rejected and so, better predicting abilities of the WV-ARCH model are confirmed for all the datasets. Indeed, differences in terms of mean of QLIKE loss functions are significant. Among all other models, GARCH-oriented models have better results

realized library. The realized volatility is estimated using a Kernel method [1].

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

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

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

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

Table 5: QLIKE losses and DMW statistics for the three series of log returns. *, ** and *** respectively signify rejecting the null hypothesis of equal losses for 90%, 95% and 99% confidence levels. Forecasts are done during 400 trading days, between the 27th of August 2007 and the 6th of March 2009. The benchmark is the WV-ARCH model.

than ARCH and NP-ARCH. This can be explained by the different ways we include the volatility persistence in GARCH- (β, as in Table 3) and ARCH-oriented models (only implied by the rolling-window procedure). However, when comparing QLIKE losses, differences between GARCH and GJR-GARCH do not seem significant.

4 Conclusion

In this paper we introduced a new method to model asset log-returns and volatility. Instead of using parametric and non-parametric techniques from econometrics, our approach is based on variational calculus.

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

A Estimation algorithm

For this section, we do not use the econometric subscript anymore: $y(t)$ replaces y_t. It will allow a greater clarity since subscripts are used here for other purposes, such as indexing the iteration in the estimation. The parenthesis choice is also consistent with the use in functional analysis, where wavelets and variational problems come from.

A.1 Overview of the algorithm

The estimation of x and g is based on an iterative algorithm, since both the estimations require distinct techniques. However, a similar transformation of the data is used in each iterations. Therefore, it can be extracted from the iterative loop and it can be executed only once. It must be considered as a preliminary step of the algorithm. This step relates to the wavelet approach for estimating x. Besides, the estimation of g is based on a variational approach, which supposes the computation of a line integral over ranked innovations. Since the innovations change after each estimate of x, their ranking also changes at each iteration. It is thus not a preliminary step of the algorithm. However, we present this ranking technique apart so as not to overload the global presentation of the iteration with such a technical specificity.

A.1.1 Preliminary step of the algorithm

The 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 scale 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:

$$\langle y, \psi_{j,k}\rangle = \sum_{t=0}^{T} y(t)\psi_{j,k}(t). \tag{7}$$

In fact, we decompose the signal in gross structure and details. Details are given by wavelet coefficients for a unique scale parameter j. In our examples, j is set to 4. The gross structure is given by scaling coefficients at the same scale j. They are given by:

$$\langle y, \phi_{j,k}\rangle = \sum_{t=0}^{T} y(t)\phi_{j,k}(t),$$

where ϕ is the scaling function related to the wavelet function. Further details on wavelets and its use to denoise time series can be found in [26, 15]. For the applicative part of the paper, we used a Daubechies wavelet with 4 vanishing moments.

A.1.2 Permutation of the innovations

This step of the algorithm relates to the variational approach and is repeated just before each iteration of the estimation of g. In the variational method, we will minimize an integral of a function in which g appears. When g is multidimensional, that is when $l > 1$, 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 [15]. 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 $\mathcal{E} \circ \theta$ along which the integral of g is empirically calculated. The idea is to minimize the Euclidean distance between $\mathcal{E}(\theta(t))$ and $\mathcal{E}(\theta(t+1))$ for all $t \in \{0, ..., T-1\}$. For example, if $l = 1$, we choose θ so that the innovations are sorted: $\mathcal{E}(\theta(0)) \leq \mathcal{E}(\theta(1)) \leq ... \leq \mathcal{E}(\theta(T))$. For higher l, 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 $l > 1$. 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 $l = 1$.

A.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 [10]. Indeed, $M/0.6745$ is a robust estimator for the Gaussian 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 this 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. It means that 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 $\Lambda_{i,j,k} = \lambda_i \sqrt{\langle (g_i \circ \mathcal{E}_{i-1})^2, \psi_{j,k}^2\rangle}$ where λ_i is a parameter and $\mathcal{E}_{i-1}$ is the $(i-1)$-th estimate of $\mathcal{E}$.[7] Examples indeed show that a level-dependent threshold much better performs than a constant threshold [17]. 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* [35, 11, 26]. The estimate of the reconstruction error is

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

where

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

because $\langle (g_i \circ \mathcal{E}_{i-1})^2, \psi_{j,k}^2\rangle$ is the estimated variance of the empirical wavelet coefficient $\langle y, \psi_{j,k}\rangle$ [16]. Conditionally to y, $\bar{\mathcal{S}}_i$ is an unbiased estimate of the reconstruction error. Any basic optimization algorithm enables 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:

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

[7] More precisely, for any time t and for $i \geq 1$, $\mathcal{E}_{i-1}(t) = (y(t-1) - x_{i-1}(t-1), ..., y(t-l) - x_{i-1}(t-l))$. For the estimate x_i, we use the threshold $\Lambda_{i,j,k}$, which thus depends on the previous estimate x_{i-1}. However, for $i = 0$, x_{i-1} is not defined and thus cannot be used in estimating $\mathcal{E}$. But since g_0 is a constant function, $\Lambda_{0,j,k}$ will be the same, whatever the choice made for $\mathcal{E}_{-1}$. As a consequence, $\Lambda_{0,j,k}$ is not level-dependent.

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} along a given path θ_i, which is re-estimated at each iteration.[8] 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 $g_{i+1} \circ \mathcal{E}_i$ of $g \circ \mathcal{E}$, where we introduce $\mathcal{G}_{i+1}$ which we define[9] by $\mathcal{G}_{i+1} = g_{i+1} \circ \mathcal{E}_i \circ \theta_i$:

$$\mu \frac{(\mathcal{G}_{i+1})^2 - (y \circ \theta_i - x_i \circ \theta_i)^2}{(\mathcal{G}_{i+1})^3} - \frac{d^2}{dt^2}\mathcal{G}_{i+1} = 0, \tag{8}$$

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.5. Then, in order to solve numerically this equation, we use a dynamical version of it which is expected, like in [2], to lead to a steady state after some iterations of the series of estimators $(\mathcal{G}_{i+1,n})_n$ of $\mathcal{G}_{i+1}$:

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

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:

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

$\mathcal{G}_{i+1,n}$ 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.

A.3 Improving the algorithm

Some refinements, concerning the initial condition $\mathcal{G}_{i+1,0}$ given i or the number of iterations N used to lead to the estimate $\mathcal{G}_{i+1}$, can be made in order to improve the algorithm, even though the standard conditions provided in the previous paragraph in general lead to satisfying results.

[8] More details about choosing θ_i are provided in Appendix A.1.2.

[9] $\mathcal{G}_{i+1}$ 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 $\theta_i : \{0, ..., T\} \to \{0, ..., T\}$ with $\mathcal{E}_i : \{0, ..., T\} \to \mathbb{R}^h$ and $g_{i+1} : \mathbb{R}^h \to \mathbb{R}$.

The main motivation for modifying these conditions is the fact that the choice of δ has an impact on the way the series of estimators $(\mathcal{G}_{i+1,n}(t))_n$ evolves and finally on the accuracy of the estimated news impact curve. The choice of δ or of N can hence be optimized so that the innovations $(z_t)_t$ better fit a unit Gaussian distribution.

Besides, some specific financial conditions may need a particular processing. For example, when different volatility regimes appear, one may prefer to initiate the series $(\mathcal{G}_{i+1,n}(t))_n$ with the constant function $\mathcal{G}_{i+1,0}$ estimated on some quantile of the residuals rather than on their median.

A.4 Proof of Proposition 1

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

$$y(t) - x(t) = g(\mathcal{E}(t))z_t.$$

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

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

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

$$\delta_z\left(\frac{y(t) - x(t)}{g(\mathcal{E}(t))}\right)\frac{1}{g(\mathcal{E}(t))} = \delta_{y(t)-x(t)|g(\mathcal{E}(t))}(y(t) - x(t)|g(\mathcal{E}(t))). \tag{9}$$

Thus, in a time $\theta(t)$, we get:

$$\tilde{\mathcal{L}}(t, \mathcal{G}(t)) = \log(g(\mathcal{E}(\theta(t)))) + \frac{1}{2}\left(\frac{y(\theta(t)) - x(\theta(t))}{g(\mathcal{E}(\theta(t)))}\right)^2.$$

A.5 Justification of equation (8)

The maximum-likelihood problem consists in maximizing the right-hand side of equation (9). 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:

$$\log(g(\mathcal{E}(t))) + \frac{1}{2}\left(\frac{y(t) - x(t)}{g(\mathcal{E}(t))}\right)^2.$$

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

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

where $\mathcal{G} = g \circ \mathcal{E} \circ \theta$. 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:

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

where

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

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

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

References

[1] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003), *Modeling and forecasting realized volatility*, Econometrica, 71, 2: 529-626

[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] Badescu, A., Elliott, R.J. and Ortega, J.P. (2015), *Non-Gaussian GARCH option pricing models and their diffusion limits*, European journal of operational research, 247, 3: 820-830

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

[5] Bollerslev, T. (1986), *Generalized autoregressive conditional hHeteroskedasticity*, Journal of econometrics, 31, 3: 307-327

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

[7] Corradi, V. (2000), *Reconsidering the continuous time limit of the GARCH(1, 1) process*, Journal of econometrics, 96, 1: 145-153

[8] Deelstra, G., Grasselli, M. and Koehl, P.F. (2004), *Optimal design of the guarantee for defined contribution funds*, Journal of economic dynamics and control, 28, 11: 2239-2260

[9] Diebold, F.X. and Mariano, R.S. (2002), *Comparing predictive accuracy*, Journal of business & economic statistics, 20, 1: 134-144

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

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

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

[13] Engle, R.F. and Ng, V.K. (1993), *Measuring and testing the impact of news on volatility*, The Journal of finance, 48, 5: 1749-1778

[14] Fiorentini, G., Calzolari, G. and Panattoni, L.(1996), *Analytic derivatives and the computation of GARCH estimates*, Journal of applied econometrics, 11, 4: 399-417

[15] Garcin, M. (2016), *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*, CRC/Taylor & Francis

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

[17] Garcin, M. and Guégan, D. (2016), *Wavelet shrinkage of a noisy dynamical system with non-linear noise impact*, Physica D: nonlinear phenomena, 325: 126-145

[18] Giaquinta, M. and Hildebrandt, S. (1996), *Calculus of variations. A series of comprehensive studies in mathematics*, Springer, Berlin, Germany

[19] 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

[20] Gouriéroux, C. and Monfort, A. (1992), *Qualitative threshold ARCH models*, Journal of econometrics, 52, 1:159-199

[21] Härdle, W. and Tsybakov, A. (1997), *Local polynomial estimators of the volatility function in nonparametric autoregression*, Journal of econometrics, 81, 1: 223-242

[22] Hafner, C.M. and Linton, O. (2010), *Efficient estimation of a multivariate multiplicative volatility model*, Journal of econometrics, 159, 1: 55-73

[23] Han, H. and Kristensen, D. (2014), *Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates*, Journal of business & economic statistics, 32, 3: 416-429

[24] Han, H. and Kristensen, D. (2015), *Semiparametric multiplicative GARCH-X model: Adopting economic variables To explain volatility*, Working paper

[25] Linton, O. and Mammen, E. (2005), *Estimating semiparametric ARCH(∞) models by kernel smoothing methods*, Econometrica, 73, 3: 771-836

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

[27] Malliavin, P. and Thalmaier, A. (2006), *Stochastic calculus of variations in mathematical finance*, Springer, Berlin, Germany

[28] Mandelbrot, B. (1963), *The variation of certain speculative prices*, The Journal of business, 36, 1: 392-417

[29] Nelson, D.B. (1990), *ARCH models as diffusion approximations*, Journal of econometrics, 45, 1: 7-38

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

[31] Newey, W.K. and West, K.D. (1986), *A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix*, Econometrica 55, 3: 703-ĂŞ708

[32] Pagan, A.R. and Schwert, G.W. (1990), *Alternative models for conditional stock volatility*, Journal of econometrics, 45, 1: 267-290

[33] Patton, A.J. and Sheppard, K. (2009), *Evaluating volatility and correlation forecasts*, in Mikosch, Th., Kreiss, J.-P., Davis, R. and Andersen, T.G. (editors), *Handbook of financial time series*, Springer, Berlin, Germany: 801-838

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

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

[36] Weitzmann, M. (2009), *Income, wealth, and the maximum principle*, Harvard university press

[37] West, K.D. (1996) *Asymptotic inference about predictive ability*, Econometrica, 64, 5: 1067-1084

[38] Zheng, Z., Qiao, Z., Takaishi, T., Stanley, H.E. and Li, B. (2014), *Realized volatility and absolute return volatility: a comparison indicating market risk*, PloS one, 9, 7: doi 10.1371