# Description, modeling and forecasting of data with optimal wavelets

Oriol Pont, Antonio Turiel, Conrad J. Perez-Vicente

# Description, modelling and forecasting of data with optimal wavelets

**Oriol Pont · Antonio Turiel ·
Conrad J. Perez-Vicente**

**Abstract**    Cascade processes have been used to model many different self-similar systems, as they are able to accurately describe most of their global statistical properties. The so-called optimal wavelet basis allows to achieve a geometrical representation of the cascade process-named microcanonical cascade- that describes the behavior of local quantities and thus it helps to reveal the underlying dynamics of the system. In this context, we study the benefits of using the optimal wavelet in contrast to other wavelets when used to define cascade variables, and we provide an optimality degree estimator that is appropriate to determine the closest-to-optimal wavelet in real data. Particularizing the analysis to stock market series, we show that they can be represented by microcanonical cascades in both the logarithm of the price and the volatility. Also, as a promising application in forecasting, we derive the distribution of the value of next point of the series conditioned to the knowledge of past points and the cascade structure, i.e., the stochastic kernel of the cascade process.

**Keywords**    Optimal wavelet · Cascade processes · Microcanonical multifractal formalism · Time series forecasting

O. Pont (✉)
Departament de Física Fonamental, Universitat de Barcelona,
Diagonal, 647, 08028 Barcelona, Spain
e-mail: oriol@ffn.ub.es

C. J. Perez-Vicente
e-mail: conrad@ffn.ub.es

A. Turiel
Physical Oceanography Department, Institut de Ciències del Mar, CSIC,
Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Spain
e-mail: turiel@icm.csic.es

## 1 Introduction

Stock markets, as well as many other trading markets, are formed by a great diversity of interacting agents, each with their own characteristics, such as reacting times, budgetary constraints and so on. As a consequence, and due to the large amount of agents taking part in a typical market, many econometric indicators behave in a complex, scale-invariant fashion, a feature that has been taken into account in many different models (Arneodo et al. 1998; Calvet and Fisher 2001; Perelló et al. 2004). However, scale invariance can be exploited not only in the design of models, but also in analysis tools capable of extracting new information from time series of dynamical systems. One of the most promising theories for the description of scale-invariant data concerns multifractal systems (Muzy et al. 2001; Turiel and Pérez-Vicente 2003), and more particularly multifractal systems in the microcanonical approach (Turiel and Pérez-Vicente 2003, 2005; Turiel et al. 2008b). A full discussion of the microcanonical multifractal formalism (MMF) exceeds the scope of this paper; here we will just make use of those facts relevant for our study, namely the existence of adequate wavelet representations for which data can be represented as an explicit cascade process across scales. With the aid of *microcanonical cascades* one can maximize the amount of information that some scales convey about the others. Besides, it is possible to provide an analytical model describing the evolution of the series, and produce high quality forecasts for both the returns and volatility evolution of stock market data.

The paper is structured as follows: next section introduces the data that will be used to illustrate the theory. Then, in Sect. 3 the multiscale cascade model is introduced. In Sect. 4 the scale invariant properties of cascade are discussed, and Sect. 5 is devoted to introduce the optimal wavelet, for which the cascade is directly verified; as an application to econometric data, the best wavelet from a bank of standard wavelet bases is obtained. Then, the direct cascade model furnished by the optimal wavelet is used to derive analytically the stochastic kernel, namely the distribution of future values conditioned by the known past values, and the entire Sect. 6 is devoted to this goal. Finally, we present in Sect. 7 our conclusions.

## 2 Description of the data and notation

We have processed a group of data belonging to the Spanish stock market (IBEX). The group is formed by daily series of 35 different assets (those with the largest liquidity in the Spanish market) during approximately ten years (from June 1996 to June 2006, although some series are slightly shorter) containing 76,663 points. We do not try to correct systematic deviations by any mean. In that sense, we always identify the ending of a session as the instant just preceding the opening of the following, no matter the actual time interval between them (sometimes several non-working days). An example of this series is shown in Fig. 1.

We are interested in relative variations of the price, i.e., the ratio of the absolute value to the absolute variation. For that reason, we will work on series formed by logarithms of prices. In this way the absolute variation between two consecutive instants
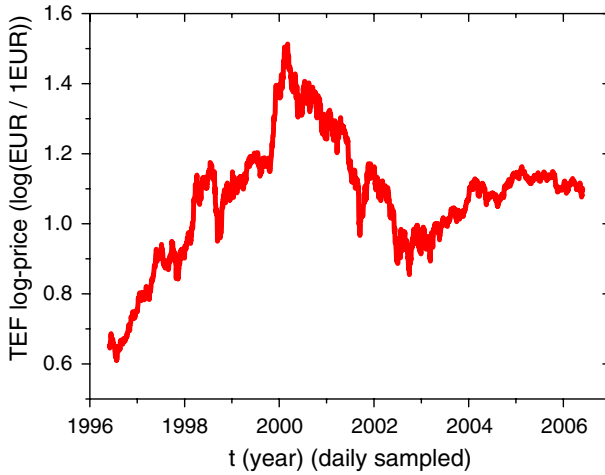
**Fig. 1** Daily series of the logarithm of quotation price for Telefónica (TEF) between June 3rd 1996 and May 30th 2006

(approximately, the derivative with respect to time) approximates the relative variation for the original stock series.

These series have been shown to have multifractal properties in the sense of MMF (Turiel and Pérez-Vicente 2003, 2005; Turiel et al. 2008b), so they are appropriate for the present study. Throughout the paper, time series will be represented by a discrete, time-ordered collection of values $x_1, x_2, \ldots$, where $x_n = x(t_n)$ and the sampling times $t_n$ are equally spaced, $t_i = t_0 + n\Delta t$. For description and forecast, we will take a finite number $N$ of past, known values, forming a vector $\mathbf{x} = (x_1, \ldots, x_N)$. The unknown value at the following time instant (which would correspond to $x_{N+1}$) will be denoted with a different letter, $y$, to emphasize that this is the forecast.

## 3 Wavelet representation of cascade processes

Scale-invariant systems cannot be described appropriately by any statistical quantity covering only one or few scales; on the contrary, to gain full understanding of those systems the influence of all scales must be taken into account. Usual data representations are based on a single observation scale, the most natural one being the rate at which the series has been sampled. For a scale-invariant signal to focus exclusively on a single scale leads to an inefficient information processing, which motivates the use of other representations where the impact of the different scales can be explicitly decomposed. This is the case of wavelet bases (Daubechies 1992; Mallat 1999). For the sake of simplicity, we will just consider orthonormal, dyadic wavelet bases.

Given a discrete or continuous time series $x(t)$, we can represent it in a dyadic wavelet basis of a mother wavelet $\Psi$ as follows:

$$x(t) = \sum_{j=-\infty}^{j_0} \sum_k \alpha_{j,k} \Psi_{j,k}(t) + \sum_k \beta_{j_0,k} \Phi_{j_0,k}(t) \tag{1}$$

where $\Phi$ is the unit function associated to $\Psi$ and $\Psi_{j,k}(t) = 2^{-j/2}\Psi\,(2^{-j}t - k)$ (Mallat 1999). In this representation, the first term accounts for the details, i.e., the variations that contribute at different scales, starting from the infinitely resolved one ($j = -\infty$) and including all of the coarser scales up to a conventionally fixed one ($j_0$). The second term represents the approximation at that scale $j_0$, which accounts for all of the remaining variations smoother than that of $j_0$. The wavelet coefficients $\alpha_{j,k}$ and the approximation coefficients $\beta_{j_0 k}$ can be easily obtained when the wavelet basis is orthonormal by simple projection:

$$\alpha_{j,k} = \langle \Psi_{j,k}, x \rangle = \int dt\, \Psi_{j,k}(t)\, x(t) \tag{2}$$

and

$$\beta_{j,k} = \langle \Phi_{j,k}, x \rangle = \int dt\, \Phi_{j,k}(t)\, x(t) \tag{3}$$

For a signal defined on an unbounded domain, the separation scale $j_0$ is chosen conventionally (eventually $j = +\infty$, what makes the second contribution to vanish). A different situation arises when the signal has a bounded domain. First, all the sums on the position $k$ are limited to positions included in the domain. Second, $j_0$ cannot be greater than the integral scale $J$ ($N = 2^J$), and for that reason it is normally chosen as it. On the other hand, if the signal is discretised, this imposes a minimum possible scale (so truncating the lower end of the sum in the first term) which can conventionally set to $j = 0$. Hence, in practice Eq. (1) is reduced to:

$$x_n = \sum_{j=0}^{J} \sum_{k \in \mathbb{Z}_{2^j}} \alpha_{j,k}\, \Psi_{j,k}(t_n) + \beta_{J,0}\, \Phi_{J,0}(t_n) \tag{4}$$

Notice that as wavelets form a basis, the wavelet coefficients in Eqs. (1) and (4) are algebraically independent. But when signals have a multifractal structure there is an additional constraint evidenced by wavelet coefficients. This is the cascade relation (Turiel and Parga 2000b; Turiel et al. 2003; Pont et al. 2008), which relates a wavelet coefficient at a coarse scale (the *Parent coefficient*) with the wavelet coefficients at the next finer scale and with the same position at the coarse level (the *Children coefficients*). The cascade relation is expressed as follows:

$$\alpha_{j,k} \doteq \eta_{j,k}\, \alpha_{j-1,\left[\frac{k}{2}\right]} \tag{5}$$

where the symbol "$\doteq$" means that both sides are equally distributed and the variables $\eta_{j,k}$ are equally distributed for any $j$, $k$, and are independent from $\alpha_{j-1,\left[\frac{k}{2}\right]}$. The variables $\eta_{j,k}$ are known as cascade variables and are related to multifractal invariants in the MMF (Turiel et al. 2008b; Pont et al. 2008). The notation $\left[\frac{k}{2}\right]$ means the integer part, rounding down, of $\frac{k}{2}$; this implies that one parent coefficient has exactly two

children coefficients (e.g. $k = 2k'$ and $k = 2k' + 1$). Notice that for any wavelet we could always define the child/parent ratio $\tilde{\eta}_{j,k}$, namely:

$$\tilde{\eta}_{j,k} = \frac{\alpha_{j,k}}{\alpha_{j-1,\left[\frac{k}{2}\right]}} \tag{6}$$

In general $\tilde{\eta}_{j,k}$ is not independent of $\alpha_{j-1,\left[\frac{k}{2}\right]}$ except for a particular choice of wavelet basis, called optimal wavelet basis, for which the child/parent ratio $\tilde{\eta}_{j,k}$ is independent of the parent coefficient $\alpha_{j-1,\left[\frac{k}{2}\right]}$ and hence it can be considered a cascade variable. This kind of wavelet is discussed in Sect. 5.

## 4 Scale invariant properties of the cascade processes

A dyadic cascade variable, i.e., a variable verifying Eq. (5), cannot follow an arbitrary distribution. In fact, the variables $\eta_{j,k}$ must follow an infinitely divisible distribution (Novikov 1994; She and Leveque 1994; Turiel and Parga 2000a), for a scale ratio of $\frac{1}{2}$. An infinitely divisible distribution implies that the product of two cascade variables is also a cascade variable, with a scale ratio equal to the product of the scale ratios (Turiel and Parga 2000a). Cascade variables $\eta$ are not scale invariant, and neither is their distribution. However, they can be related to scale invariant quantities through the introduction of the MMF (Turiel et al. 2008b). We will not present a full discussion of MMF here, just a simplified, more qualitative introduction which would serve to support our experiences.

According to Eq. (5), any wavelet coefficient $\alpha_{j,k}$ is distributed as the product of $j$ independent dyadic cascade variables, in the way:

$$\alpha_{j,k} \doteq \prod_{j'=0}^{j-1} \eta_{j-j',\left[\frac{k}{2^{j'}}\right]} \alpha_{0,0} \tag{7}$$

If the variables $\eta_{j,k}$ are infinitely divisible, they verify:

$$\eta_{j,k} = \left(\frac{1}{2}\right)^{\tilde{h}_{j,k}} \tag{8}$$

where $\tilde{h}_{j,k}$ is a scale-invariant quantity, the transition singularity exponent between the scales $j$ and $j-1$ at the point $k$ (Pont et al. 2008; Turiel et al. 2008b). We can thus define a punctual estimate of the singularity exponent associated to the whole cascade at the point $2^{-j}k$ by taking logarithms in Eq. (7) and normalizing by the scale factor $j \log 2$, namely:

$$\hat{h}_{j,k} = -\frac{\log \alpha_{j,k}}{j \log 2} \doteq h_{j,k} - \frac{\log \alpha_{0,0}}{j \log 2} \tag{9}$$

where $h_{j,k}$ is a dimensionless, scale invariant field known as singularity exponent, which is the average of the transition singularity exponents,

$$h_{j,k} = \frac{1}{j} \sum_{j'=0}^{j-1} \tilde{h}_{j-j', \left[\frac{k}{2^{j'}}\right]} \tag{10}$$

Singularity exponents characterize the degree of regularity or irregularity of a function at a point, and serve to characterize the multifractal geometry of scale invariant fields in MMF (Turiel et al. 2008b). The multifractal structure of the signal is intimately linked to its dynamic properties (Turiel and Pérez-Vicente 2003, 2005; Turiel et al. 2005a,b, 2008a). This fact explains why it is so much important to obtain the singularity exponents for tasks such as reconstruction or forecasting. Punctual estimates as the one given by Eq. (9) are good approximations to the actual value of the singularity exponent when the scale ratio is large enough, namely $j$ is large enough and so the second factor in Eq. (9) can be neglected (Turiel et al. 2006; Pont et al. 2006).

The distribution of singularity exponents as obtained at a scale $j$ is not scale invariant. However, it can be related to the singularity spectrum of the underlying multifractal hierarchy (Frisch 1995; Arneodo et al. 1995). Let $r = 2^{-j}$ be the scale ratio associated to $h_{j,k}$, according to Eq. (10). The distribution of values of $h_{j,k}$, $\rho(h_{j,k})$ verifies:

$$\rho(h_{j,k}) = A_0 r^{d-D(h_{j,k})} \tag{11}$$

where $d$ is the dimension of the embedding space ($d = 1$ for time series) and $D(h)$ is the function relating the value of the singularity exponent, $h$, with the fractal dimension of the associated singularity component (Falconer 1990; Turiel et al. 2008b). If the mode (i.e., most probable value) of $h_{j,k}$ has a known dimension [which must be $d$ if the support of the multifractal is the whole space (Turiel et al. 2006, 2008b)], then the singularity spectrum can be directly retrieved from Eq. (11) by a log–log transformation (Turiel et al. 2006):

$$D(h_{j,k}) = d + \frac{\rho(h_{j,k})/\rho_0}{j \log 2} \tag{12}$$

where $\rho_0$ is the maximum value of the function $\rho(h_{j,k})$. The singularity spectrum is a global scale invariant quantity, so for any value $h_0$ and any $j, k, j', k'$ we should observe $D(h_{j,k} = h_0) = D(h_{j'k'} = h_0)$. In fact, the equality of these factors at different scales is one of the conditions for MMF to be valid (Turiel et al. 2008b). In Fig. 2 we show that the experimental singularity spectra obtained from the series of returns and from the series of volatilities (derived from our IBEX 35 dataset) are coincident within the experimental uncertainty at three different scales [notice that the right tail is always worse determined; see discussion in Turiel et al. (2006)]. Remarkably enough, return and volatility singularity spectra are very similar. This should be expected because volatility is a measure of the amplitude of returns. As singularity exponents describe function regularity (Turiel et al. 2008b), those of returns approximately coincide with those of volatilities.
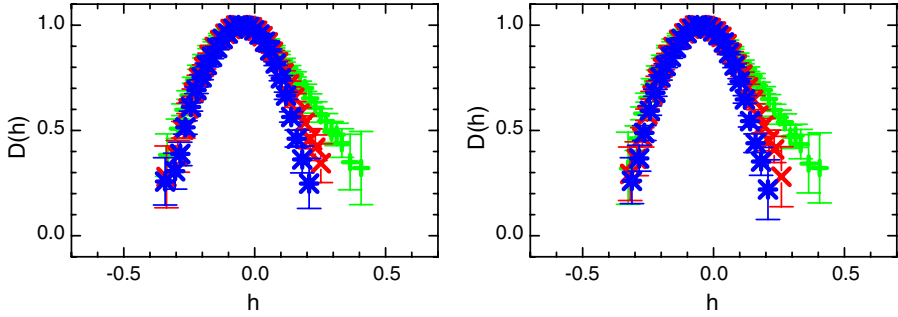
**Fig. 2** Singularity spectra derived from returns (*left*) and volatilities (*right*). The absolute values of returns and the squares of volatilities have been averaged at three different time windows, i.e., they have been projected with a box wavelet. The scales (sizes of the windows) are as follows: +: 2 days; ×: 4 days; *:8 days. For simplicity, we use a box wavelet, as we have seen that the marginal distributions of singularity exponents, $\rho_{h_{j,k}}$, are the same with almost any wavelet (the optimal wavelet basis is relevant only for calculation of the parent/child joint distribution)

Let us finally remark that the knowledge of the singularity spectrum $D(h)$ provides information of the distribution of the cascade variables at any scale. Simply, by the application of Eqs. (9) and (11) we can know the distribution of $\alpha_{j,k}$ at any scale $j$.

## 5 Optimal wavelet

### 5.1 Optimality and the $Q$-parameter

As commented before, an optimal wavelet is such that the variable $\tilde{\eta}_{j,k}$ is optimal, namely it is independent of the parent wavelet coefficient $\alpha_{j-1,\left[\frac{k}{2}\right]}$ and is distributed according a multiplicative cascade. This choice of wavelet basis is very convenient, as it implies extracting the independent factors relating scales thus minimizing redundancy, what can be used for tasks such as coding or inference. The existence of an optimal wavelet basis was first hypothesized in (Turiel and Parga 2000b) for the case of natural images, and in the same reference a constructive formula able to retrieve the optimal wavelet from a set of data was given. However, that formula requires to have a very large training dataset with many independent members, something which is rarely the case. For that reason, some attempts to design less data-demanding algorithms to retrieve the optimal wavelet from data have recently been enterprised. In Pont et al. (2008) a criterion to assess the degree of optimality of a wavelet basis was presented. It is based on a parameter named $Q$, which compares the expectation value of the ratio child/parent with the expectation value of a cascade variable (which is a priori known from some constraints). More precisely, $Q$ is defined as follows:

$$Q = \frac{\langle|\tilde{\eta}|\rangle}{\langle|\eta|\rangle} \tag{13}$$

where the expectation values $\langle\cdot\rangle$ are taken on an ensemble of realizations (including different scales $j$ and positions $k$). A wavelet is optimal if and only if $Q = 1$, and

any non-optimal wavelet will have $Q > 1$ (Pont et al. 2008). The $Q$ parameter has many advantages with respect to other criteria on wavelet optimality, as for instance the parameter $I$ measuring the mutual information between $\tilde{\eta}$ and its parent coefficient ($I \geq 0$ and $I = 0$ if and only if the optimal wavelet). First, $Q$ varies continuously under perturbations on the optimal wavelet, what makes it more appropriate for continuous strategies of optimization. Second, $Q$ is less demanding in data and more stable, giving a more accurate description of the wavelet properties than $I$.

Another, more visual criterion to assess the optimality of a wavelet basis consists in plotting joint histograms of $\log |\alpha_C|$ versus $\log |\alpha_P|$ (where the subindexes $C$ and $P$ denote "Children" and "Parent", respectively, used here to alleviate notation). This kind of histogram plots were introduced by Simoncelli and co-workers (Buccigrossi and Simoncelli 1999; Wainwright et al. 2001). If the wavelet is optimal, the equality in Eq. 5 implies that:

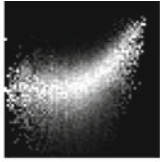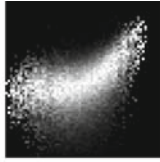$$\log |\alpha_C| = \log |\eta| + \log |\alpha_P| \tag{14}$$
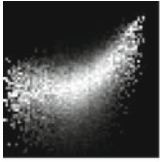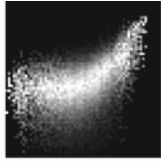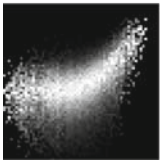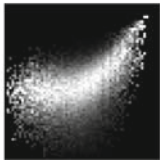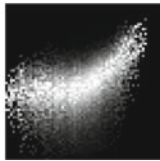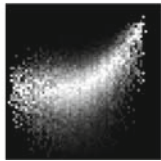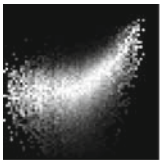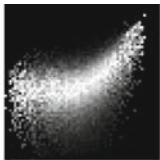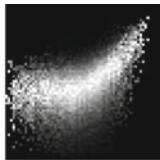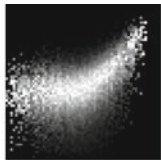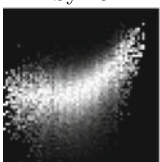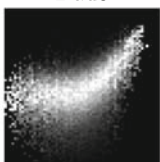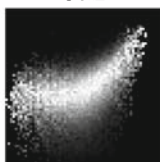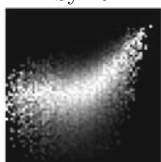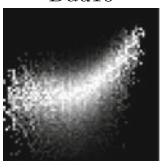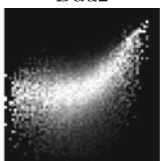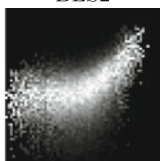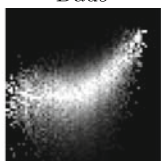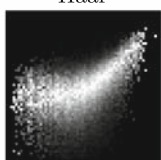
Hence, the joint histogram of $\log |\alpha_C|$ versus $\log |\alpha_P|$ for an optimal wavelet must have a shape of straight line with slope 1, with the values of $\log |\eta|$ distributed around it. If the wavelet is not optimal, for small values of $\log |\alpha_P|$ a deviation from this linear behavior will be observed and the predictability of $\alpha_P$ about $\alpha_C$ diminishes. As a consequence, $\alpha_C$ becomes independent of $\alpha_P$ and the histogram displays a horizontal straight line. It is thus possible to roughly assess the degree of optimality of a wavelet by knowing the size of the range on which the optimal linear behavior, Eq. 14, holds. In general, however, it is better to work with more quantitative measurements such as $Q$, using joint histograms just for illustration purposes. We show examples of joint histograms derived from real data in the following.

### 5.2 Optimal wavelet for Spanish stock market series

We have constructed a bank of 24 orthonormal dyadic wavelet bases. The bases included in this bank are some of the most frequently used in the wavelet literature: Haar, Daubechies of orders 2–10 (signatures Dau2-9 and DauA), Symlets of orders 4–9 (signatures Sym4-9), Coiflets of orders 1 to 5 (signatures Coi1-5), and spline or Battle-Lemarié of orders 1,2,3 and 6 (signatures BL1-3 and BL6). We have searched which wavelet in this bank is closest to optimality for our dataset according to the $Q$ criterion. Results are summarized in Table 1.

It is observed that the best wavelet (in terms of optimality) in the bank is the Battle-Lemarié wavelet of order 3. A similar experience conducted on volatility series throws that the optimal w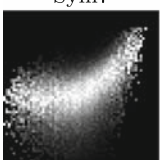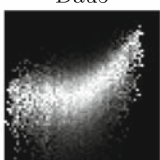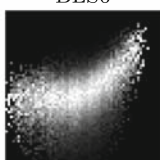avelet is a different one, Symlet of order 7 (in that case, Sym7 attains a $Q = 2.22$ while for other wavelets of the bank the average value of $Q$ is 2.95 and the farthest from optimality is Haar with $Q = 5.71$). Should one expect to have a relation between the optimal wavelet for log-prices and that of volatility? In fact the answer is yes. Volatility is the amplitude of return variations, i.e., it represents some kind of modulus of the derivative of log-prices, and the optimal wavelet of the derivative series is the derivative of the optimal wavelet of the series, as explained in

**Table 1** Histograms of child wavelet coefficient $\alpha_C$ conditioned by the value of its parent coefficient $\alpha_P$ (in logarithmic scale) and values of $Q$ for different wavelet bases

| BLS3 | Sym6 | Coi3 | Sym4 |
|------|------|------|------|
| 1.87 | 1.95 | 1.95 | 1.99 |
| Coi5 | BLS1 | Coi4 | Coi1 |
| 2.00 | 2.01 | 2.06 | 2.09 |
| Dau4 | Dau3 | Dau7 | Dau6 |
| 2.10 | 2.11 | 2.18 | 2.20 |
| Sym8 | Dau8 | Coi2 | Sym5 |
| 2.20 | 2.24 | 2.27 | 2.35 |
| Dau10 | Dau2 | BLS2 | Dau9 |
| 2.44 | 2.45 | 2.46 | 2.46 |
| Sym7 | Dau5 | BLS6 | Haar |
| 2.56 | 2.57 | 2.60 | 2.77 |

They are sorted from closest to farthest from optimality, according to the value of $Q$. All histograms range from $-6$ to $3.5$ in both axes. The elbow shape is a direct consequence of loss of optimality, so its position depends on the degree of optimality (for Battle-Lemarié 3, the best wavelet, it is located the leftmost, at about $\frac{1}{3}$ of the histogram, while for Haar, the worst wavelet, it is the rightmost, at about $\frac{1}{2}$ of the histogram)

**Fig. 3** Comparison of the
optimal wavelet obtained for
volatility series (Symlet of order
7) and the time derivative of
optimal wavelet obtained for the
logarithm of the quotation price
\(Battle-Lemarié of order 3)
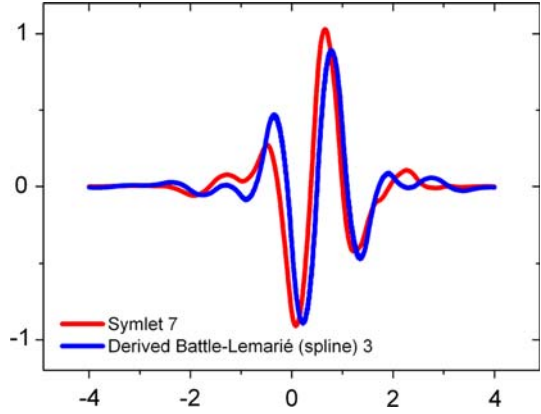


Turiel and Parga (2000b). As shown in Fig. 3, Symlet of order 7 is very similar to
the derivative of Battle-Lemarié of order 3. This confirms the validity of the cascade
description in terms of optimal wavelets for both series.

## 6 Conditioned distribution of quotation values according the cascade model

### 6.1 Settings

We will try now to determine the distribution of the series value $y$ conditioned to the
knowledge of the vector $\mathbf{x}$ of $N$ previous events, that we will denote by $\rho_y(y|\mathbf{x})$. Due
to the high dimensionality of this function, it cannot be estimated in general and it
rather needs to be modeled. We will make use of the cascade properties that we have
introduced in the previous sections to model this conditioned PDF. First, we assume
that our wavelet basis is optimal, so:

$$\alpha_{j,k} = \eta_{j,k}\, \alpha_{j-1,\left[\frac{k}{2}\right]} \tag{15}$$

where the variables $\eta_{j,k}$ are all identically distributed according to a known PDF $\rho_\eta$
and each variable $\eta_{j,k}$ is independent of $\alpha_{j-1,\left[\frac{k}{2}\right]}$. As $\alpha_{j-1,\left[\frac{k}{2}\right]}$ can be further decom-
posed as $\eta_{j-1,\left[\frac{k}{2}\right]}\alpha_{j-2,\left[\frac{k}{4}\right]}$, we conclude that $\eta_{j,k}$ is independent of its grand-parent
$\eta_{j-2,\left[\frac{k}{4}\right]}$ and by induction, it is independent of all its ancestors. Notice however that
this does not imply that $\eta_{j,k}$ is independent of $\eta_{j,k'}$; the horizontal correlations must
be studied and implemented.

Due to the linearity in the definition of the wavelet coefficients, Eq. (2), $\alpha_{j,k}$ can
be expressed as a linear function of $y$, with coefficients depending on $\mathbf{x}$, namely:

$$\alpha_{j,k} = a_{j,k}\, y + b_{j,k} \tag{16}$$

Hence, the cascade variables are expressed as:

$$\eta_{j,k} = \frac{a_{j,k}\, y + b_{j,k}}{a_{j-1,\left[\frac{k}{2}\right]}\, y + b_{j-1,\left[\frac{k}{2}\right]}} \tag{17}$$

As there is a deterministic relation between each cascade variable and the variable $y$, the conditioned distribution of the cascade variables is expressed as a combination of delta functions, namely:

$$\rho\left(\{\eta_{j,k}\}|y, \mathbf{x}\right) = \prod_{j,k} \delta\left(\eta_{j,k} - \frac{a_{j,k}\, y + b_{j,k}}{a_{j-1,\left[\frac{k}{2}\right]}\, y + b_{j-1,\left[\frac{k}{2}\right]}}\right) \tag{18}$$

If we integrate this distribution with $\rho_y(y|\mathbf{x})$ we will obtain the distribution of cascade variables conditioned by $\mathbf{x}$ only, namely:

$$\rho\left(\{\eta_{j,k}\}|\mathbf{x}\right) = \int \mathrm{d}y\, \rho_y(y|\mathbf{x})\, \rho(\{\eta_{j,k}\}|y, \mathbf{x}) \tag{19}$$

We need to propose a model for $\rho\left(\{\eta_{j,k}\}|\mathbf{x}\right)$ so we can solve for $\rho_y(y|\mathbf{x})$.

## 6.2 The model

We propose the following model for $\rho_y(y|\mathbf{x})$:

$$\rho\left(\{\eta_{j,k}\}|\mathbf{x}\right) = \kappa \left(\prod_{j,k} \rho_\eta\left(\eta_{,k}\right)\right) \chi_{\mathbf{x}}(\{\eta_{j,k}\}) \tag{20}$$

where $\chi_{\mathbf{x}}(\{\eta_{j,k}\})$ is a set function (it takes only the values 0 and 1) that restricts the values of the variables $\eta_{j,k}$ to a subset of really accessible values. That is, we assume that the probabilities are independent but not every possible value of $\eta_{j,k}$ will be visited once $\mathbf{x}$ is fixed, which explains the normalization prefactor $\kappa$. Nevertheless, we assume that the pyramid is large enough not to modify the shape of the marginal distributions.

## 6.3 The solution

Recalling Eq. (19) and using Eq. (18) we have:

$$\rho\left(\{\eta_{j,k}\}|\mathbf{x}\right) = \int \mathrm{d}y\, \rho_y(y|\mathbf{x}) \prod_{j,k} \delta\left(\eta_{j,k} - \frac{a_{j,k}\, y + b_{j,k}}{a_{j-1,\left[\frac{k}{2}\right]}\, y + b_{j-1,\left[\frac{k}{2}\right]}}\right) \tag{21}$$

We will next show that the following function:

$$\rho_y(y|\mathbf{x}) = \kappa \prod_{j,k} \rho_\eta \left( \frac{a_{j,k}\, y + b_{j,k}}{a_{j-1,\left[\frac{k}{2}\right]}\, y + b_{j-1,\left[\frac{k}{2}\right]}} \right) \tag{22}$$

verifies the proposed model, Equation (20). The values of $\eta_{j,k}$ actually visited are those of the shape $\left\{ \eta_{j,k} = \frac{a_{j,k}\, y + b_{j,k}}{a_{j-1,\left[\frac{k}{2}\right]}\, y + b_{j-1,\left[\frac{k}{2}\right]}} \right\}_{j,k}$ where $y$ can vary but must be the same for all $\eta_{j,k}$ in the same realization of the pyramid. Substituting Eq. (22) in Eq. (21) we have:

$$\rho\ \{\eta_{j,k}\}|\mathbf{x}) = \kappa \int dy \prod_{j,k} \rho_\eta\ (\eta_{,k}\ ) \delta\!\left( \eta_{j,k} - \frac{a_{j,k}\, y + b_{j,k}}{a_{j-1,\left[\frac{k}{2}\right]}\, y + b_{j-1,\left[\frac{k}{2}\right]}} \right) \tag{23}$$

If we assume that the pyramid is large enough so that the distribution of values $\eta_{j,k}$ across the pyramid is not dependent on the particular value of $y$ considered, we can take the product of PDFs outside the integral and hence:

$$\rho\ \{\eta_{j,k}\}|\mathbf{x}) = \kappa \left( \prod_{j,k} \rho_\eta\ (\eta_{,k}) \right) \int dy \prod_{j,k} \delta\!\left( \eta_{j,k} - \frac{a_{j,k}\, y + b_{j,k}}{a_{j-1,\left[\frac{k}{2}\right]}\, y + b_{j-1,\left[\frac{k}{2}\right]}} \right) \tag{24}$$

Notice that the remaining integral is a set function with support on the values of $\eta_{j,k}$ which are accessible only. Hence,

$$\chi_{\mathbf{x}}(\{\eta_{j,k}\}) = \int dy \prod_{j,k} \delta\!\left( \eta_{j,k} - \frac{a_{j,k}\, y + b_{j,k}}{a_{j-1,\left[\frac{k}{2}\right]}\, y + b_{j-1,\left[\frac{k}{2}\right]}} \right) \tag{25}$$

and so Eq. (20) follows. We conclude that Eq. (22) is the expression of the distribution of $y$ conditioned by the past values $\mathbf{x}$ according to the cascade model.

## 6.4 Asymptotic limits and stability of the solution

The stochastic model given by Eq. (22) is very appealing, but its application to real data requires some care. First of all, let us realize that the model is based on the assumption that the wavelet used is optimal. This is a bit delicate for the case in which the denominators $a_{j-1,\left[\frac{k}{2}\right]}\, y + b_{j-1,\left[\frac{k}{2}\right]}$ vanish, as for these, as shown by the conditioned histograms in Table 1, the cascade model is no longer valid. In fact, for those values of $y$ making the denominator to vanish the numerator should also vanish, what implies $a_{j,k}/a_{j-1,\left[\frac{k}{2}\right]} = b_{j,k}/b_{j-1,\left[\frac{k}{2}\right]}$. Hence, such a constraint should be implemented in an appropriate numerical scheme in order to obtain stable results.

Another interesting limit is when $|y| \to \infty$. In that case, the solution collapses to a fixed value,

$$\rho_y(y|\mathbf{x}) \underset{|y| \to \infty}{\Rightarrow} \kappa \prod_{j,k} \rho_\eta \left( \frac{a_{j,k}}{a_{j-1,\left[\frac{k}{2}\right]}} \right) \qquad (26)$$

In order to define an integrable distribution, $\rho_y(y|\mathbf{x}) \underset{|y| \to \infty}{\Rightarrow} 0$, so this has two consequences. First, there exists a finite maximum value $\eta_\infty$ for the variable $\eta$, so $\rho_\eta$ $(\eta > \eta_\infty) = 0$. Second, for at least one $j, k$, we must have $a_{j,k}/a_{j-1,\left[\frac{k}{2}\right]} > \eta_\infty$. The first condition is in fact trivially verified, as discussed in (Turiel and Pérez-Vicente 2003; Turiel et al. 2008b): a physical signal has always this finite maximum. The second property implies that a well-realized cascade must have one of the ratios $a_{j,k}/a_{j-1,\left[\frac{k}{2}\right]}$ large enough. This property can be used as a control check on the validity of the cascade model to a given case. Let us finally remark that if the two conditions are fulfilled hence the range of valid values of $|y|$ is bounded, so there is a minimum and a maximum possible value of $y$.

## 6.5 Conditioned distribution and maximum likelihood forecasting

Since $\rho_y(y|\mathbf{x})$ takes into account the presence of the cascade, it should be a better predictor than just considering the marginal distribution of returns $\rho_r(r)$ centered on the last point: $\rho_{\hat{y}}(\hat{y}) = \rho_r(\hat{y} - x_N)$.[1] In Fig. 4 we show the distributions $\rho_y(y|\mathbf{x})$ and $\rho_{\hat{y}}(\hat{y})$ for a randomly chosen point of the TEF series. We can see that our modeled distribution, $\rho_y(y|\mathbf{x})$, is visibly skewed, meaning that the cascade at this point is making negative returns more likely than positive ones. We also see that this distribution is narrower than the return-derived distribution, i.e, for this point the cascade structure is implying a reduction of volatility. Similar results are seen in other points of all the series.

The presented model is not rigid in the sense that it does not forecast an exact future value, but a distribution of possible future values. This allows to forecast not only the most likely future value but also the instant volatility and higher-moment information conditioned by a given past. If we concentrate on the most probable value only, this is a maximum likelihood estimation of the future value. We have calculated these most likely future values sequentially and compared them to the actual observed values, all along the series and for all the series. Since volatility changes over time and also between series, we have calculated the difference between actual and predicted values, divided by the estimated instantaneous 1-day volatility. This quotient is around 0.85, similar for all the series (between 0.84 and 0.87). This means that the knowledge

---

[1] With the marginal distribution of returns $\rho_r(r)$ we can construct a simple predictor of $y$ from the last price $x_N$ as follows: $\hat{y} = \mathbf{x}_N + r$, with a random return $r$ chosen according to $\rho_r(r)$. $x_N$ is the last element of vector $\mathbf{x}$ and hence the element just preceding $y$. Obviously, the distribution of $\hat{y}$ is just the distribution $\rho_r$ shifted in its argument by $-x_N$: $\rho_{\hat{y}}(\hat{y}) = \rho_r(\hat{y} - x_N)$.

**Fig. 4** Distribution $\rho_y(y|\mathbf{x})$ for the TEF series on February 1st 2002 (*light/red*), compared to the marginal distribution of returns added to the last point $\rho_{\hat{y}}(\hat{y})$ (*dark/blue*). The actual logarithm of price that day was 2.71 (*circle*)
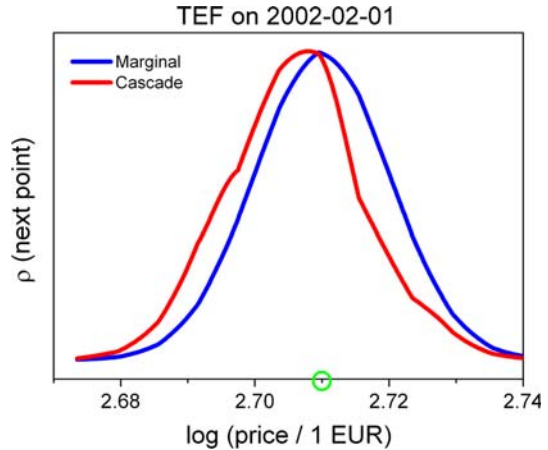


furnished by our model allows reducing expected risk in about 15% from an estimation of volatility at the same point.

## 7 Conclusions

The development of new tools designed to improve signal processing for self-similar scale-invariant systems such as multifractals is in continuous evolution. In this context, wavelets play a key role. An important concept in the description of multifractal systems is to know whether dynamical variables follow a cascade process in analogy to the case of Fully Developed Turbulence. Wavelet projections are appropriate to find and characterize such a cascade, because of their ability to tune the scale of changes in the variables. In this way, the wavelet coefficients of a signal governed by a cascade follow multiplicative relations able to characterize the cascade process and provide relevant information about its dynamics. The multiplicative factors that relate wavelet coefficients at consecutive scales are called *cascade variables* and are a representation basis of the signal. Let us remark that to deal appropriately with the cascade not any wavelet is efficient. In fact, only the optimal wavelet of the system ensures independence of the cascade variables across scales. Therefore, to find such a wavelet is extremely important due to its potential applications in encoding, forecasting and, in general, signal processing. Any other wavelet systematically induces a certain level of noise that perturbs the cascade relations. Unfortunately, the search for such a representation basis is a complex issue for real data.

In this paper we have addressed some of these important problems. First, we have shown that stock market time series can be described in terms of cascade processes. We have shown that both log-prices and volatility are multifractal observables following a multiplicative process across scales. This result reveals the existence of a hierarchical structure whose analysis can be relevant not only to improve the statistical characterization of variables but also to forecast them. To complete the analysis, we have provided a new criterion to measure the optimality degree of a wavelet in an accurate,

robust and little data-demanding way, specially appropriate for real data consisting of limited datasets. With this optimality degree estimator we have found the closest to optimal wavelet among a bank of standard wavelets for the logarithm of quotation price (Battle-Lemarié 3) and the volatility series (Symlet 7).

We have also faced explicitly the problem of forecasting. In this context, we have provided a theoretical model able to compute the probability distribution of an unknown point of the series conditioned to the knowledge of previous events making use of the cascade properties and knowing the optimal wavelet basis. As far as we know, this is the first theoretical approach in these terms and opens a new door to the problem. We have shown that in practice the distribution evolves and changes width and skewness, i.e., at some points the cascade favors positive returns while in others it favors negative returns, or similarly it implies a volatility increase at some points and decrease in others. Additionally, a simple maximum likelihood estimator shows a discrepancy smaller than the volatility at the same point.

# References

Arneodo A, Argoul F, Bacry E, Elezgaray J, Muzy JF (1995) Ondelettes, multifractales et turbulence. Diderot Editeur, Paris

Arneodo A, Muzy J-F, Sornette D (1998) "direct" causal cascade in the stock market. Eur Phys J B 2: 277–282

Buccigrossi RW, Simoncelli EP (1999) Image compression via joint statistical characterization in the wavelet domain. IEEE Trans Image Process 8:1688–1701

Calvet L, Fisher A (2001) Forecasting multifractal volatility. J Economet 105:27–58

Daubechies I (1992) Ten lectures on wavelets. CBMS-NSF Series in App. Math. Capital City Press, Montpelier

Falconer K (1990) Fractal geometry: mathematical foundations and applications. Wiley, Chichester

Frisch U (1995) Turbulence. Cambridge University Press, Cambridge

Mallat S (1999) A wavelet tour of signal processing, 2nd ed. Academic Press, New York

Muzy J-F, Sornette D, Delour J, Arneodo A (2001) Multifractal returns and hierarchical portfolio theory. Quant Finance 1:131–148

Novikov EA (1994) Infinitely divisible distributions in turbulence. Phys Rev E 50:R3303

Perelló J, Masoliver J, Bouchaud J (2004) Multiple time scales in volatility and leverage correlations: a stochastic volatility model. J Math Finance 11:27–50

Pont O, Turiel A, Pérez-Vicente C (2006) Application of the microcanonical multifractal formalism to monofractal systems. Phys Rev E 74:061110

Pont O, Turiel A, Perez-Vicente C (2008) On optimal wavelet bases for the realization of microcanonical cascade processes. Under revision in Eur Phys J B, arXiv:0805.4810v1

She ZS, Leveque E (1994) Universal scaling laws in fully developed turbulence. Phys Rev Lett 72:336–339

Turiel A, Parga N (2000a) The multi-fractal structure of contrast changes in natural images: from sharp edges to textures. Neural Comput 12:763–793

Turiel A, Parga N (2000b) Multifractal wavelet filter of natural images. Phys Rev Lett 85:3325–3328

Turiel A, Pérez-Vicente C (2003) Multifractal geometry in stock market time series. Phys A 322:629–649

Turiel A, Pérez-Vicente C (2005) Role of multifractal sources in the analysis of stock market time series. Phys A 355:475–496

Turiel A, Nadal J-P, Parga N (2003) Orientational minimal redundancy wavelets: from edge detection to perception. Vis Res 43\(9\):1061–1079

Turiel A, Grazzini J, Yahia H (2005a) Multiscale techniques for the detection of precipitation using thermal IR satellite images. IEEE Geosci Remote Sens Lett 2\(4\):447–450. doi:10.1109/LGRS.2005.852712

Turiel A, Isern-Fontanet J, García-Ladona E, Font J (2005b) Multifractal method for the instantaneous evaluation of the stream function in geophysical flows. Phys Rev Lett 95(10):104502. doi:10.1103/PhysRevLett.95.104502

Turiel A, Pérez-Vicente C, Grazzini J (2006) Numerical methods for the estimation of multifractal singularity spectra on sampled data: a comparative study. J Comput Phys  216\(1\):362–390

Turiel A, Solé J, Nieves V, Ballabrera-Poy J, García-Ladona E (2008a) Tracking oceanic currents by singularity analysis of micro-wave sea surface temperature images. Remote Sens Environ 112:2246–2260

Turiel A, Yahia H, Pérez-Vicente C (2008b) Microcanonical multifractal formalism: a geometrical approach to multifractal systems. Part I: Singularity analysis. J Phys A 41:015501

Wainwright MJ, Simoncelli EP, Willsky AS (2001) Random cascades on wavelet trees and their use in modeling and analyzing natural images. Appl Comput Harmonic Anal 11:89–123