

# Scale Dependencies and Self-Similar Models with Wavelet Scattering Spectra

Rudy Morel<sup>a,\*</sup>, Gaspar Rochette<sup>a</sup>, Roberto Leonarduzzi<sup>a</sup>, Jean-Philippe Bouchaud<sup>c</sup>, Stéphane Mallat<sup>b,a</sup>

<sup>a</sup>*École Normale Supérieure, 45 rue d'Ulm, 75005, Paris, France*

<sup>b</sup>*Collège de France, 11 place Marcelin Berthelot, 75231, Paris, France*

<sup>c</sup>*Capital Fund Management, 23 Rue de l'Université, 75007, Paris, France*

---

## Abstract

We introduce the wavelet scattering spectra which provide non-Gaussian models of time-series having stationary increments. A complex wavelet transform computes signal variations at each scale. Dependencies across scales are captured by the joint correlation across time and scales of wavelet coefficients and their modulus. This correlation matrix is nearly diagonalized by a second wavelet transform, which defines the scattering spectra. We show that this vector of moments characterizes a wide range of non-Gaussian properties of multi-scale processes. We prove that self-similar processes have scattering spectra which are scale invariant. This property can be tested statistically on a single realization and defines a class of wide-sense self-similar processes. We build maximum entropy models conditioned by scattering spectra coefficients, and generate new time-series with a microcanonical sampling algorithm. Applications are shown for highly non-Gaussian financial and turbulence time-series <sup>1 2</sup>.

*Keywords:* maximum entropy, non-Gaussianity, scattering, self-similarity, time-series, wavelets.

---



---

\*Corresponding author.

*Email address:* `rudy.morel@ens.fr` (Rudy Morel)

<sup>1</sup>This work is supported by the PRAIRIE 3IA Institute of the French ANR-19-P3IA-0001 program and the ENS-CFM models and data science chair.

<sup>2</sup>Declarations of interest: none.

## 1. Introduction

Time-series having stationary increments with variations on a wide range of scales are encountered in physics, finance, biology, medicine and many other fields. Such multi-scale time-series typically include complex intermittent phenomena with local bursts of activity, and time-asymmetries due to some form of causality. The importance of this topic was first recognized by Mandelbrot [1, 2, 3] and led to a considerable body of work on multifractal signals [4, 5, 6, 7, 8, 9, 10]. Among multi-scale processes, self-similar models have a probability distribution which is invariant to scaling, up to multiplicative factors. To validate numerically such models, it is however necessary to introduce weaker forms of self-similarity that can be estimated over limited data.

Simplified multi-scale models have been introduced from marginal distributions of signal increments, by Frisch and Parisi [11]. They define a weak form of self-similarity from a scale invariance of high order moments of these marginal distributions. This can be sufficient to detect non-Gaussian distributions. Section 2 reviews these models together with the multifractal formalism, which replaces increments by wavelet coefficients. Marginal distributions at each scale are simple to estimate, but they do not capture dependencies of signal variations across scales. These dependencies are crucial to specify many properties, in particular, the existence of transient events, which have particular signatures at multiple scales.

Models of multi-scale distributions can be defined as a maximum entropy distribution conditioned by a vector of moments  $\mathbb{E}\{\Phi(X)\}$ . If they exist, they have an exponential probability distribution

$$p_{\theta}(x) = Z_{\theta}^{-1} e^{-\theta^T \Phi(x)}.$$

for  $\theta \in \mathbb{R}^M$ , where  $M$  is the number of moments. Maximum entropy models depend only on the energy vector  $\Phi(x)$ , which needs to be chosen appropriately. Gaussian processes are maximum entropy models conditioned by first and second order moments.

A central result of the paper is the construction of  $\Phi$ , so that  $\mathbb{E}\{\Phi(X)\}$  specifies scale dependencies, and provide accurate models of multi-scale time-series. The dimension  $M$  of  $\Phi(X)$  is much smaller than the dimension of  $X$ , so that it can define a consistent mean estimator which converges to  $\mathbb{E}\{\Phi(X)\}$  when the dimension of  $X$  increases. As a result, maximum entropy models can be estimated from a single realization of  $X$ . New samples of  $X$

Wavelets used the look at how the signal behaves at different scales

They then measure how these scales interact

Combine this information into 'fingerprint' - wavelet scattering spectrum

Test self-similarity - if the fingerprint looks the same when you zoom in or out, signal behaves like a self-similar process

are generated by sampling a microcanonical model, which approximates the macrocanonical model [12].

A wavelet transform computes multi-scale signal variations. Complex wavelet coefficients carry a complex modulus and a complex phase information. Section 3 proves that wavelet coefficients are nearly uncorrelated at different scales, because their phases oscillate at different frequencies. To measure non-linear dependencies across scales, it is tempting to move towards higher order moments [13]. This requires to compute many moments with high variance estimators, which gives poor numerical results over limited size time-series. Lower variance estimators have been studied by replacing high order moments with phase harmonics [14] or by eliminating the phase with a modulus non-linearity [15, 16]. We show that scale dependencies can be captured by correlating wavelet coefficients and their modulus. We prove that self-similar processes yield normalized correlation matrices which are invariant to scaling. Section 3.3 derives a definition of wide-sense self-similarity, which is analogous to the definition of wide-sense stationarity, where invariance to translation of correlation matrices is replaced by an invariance to scaling.

Wavelet modulus cross-correlation matrices are too large to be estimated accurately from a single time-series realization. Section 4 shows that applying a second wavelet transform defines a scattering covariance matrix which is nearly diagonal. Dependencies across scales are captured by diagonal scattering cross-correlation coefficients, also called scattering cross-spectrum, which can be estimated from a single realization. We shall see that the scattering spectra provide an interpretable dashboard which captures non-Gaussian properties, including bursts of activity and time-asymmetries, as well as self-similarity.

Fractional Brownian motions, Poisson processes, multifractal random walks and Hawkes processes are often used as models of multi-scale processes which may or may not be self-similar. Section 5 shows that the scattering spectra reveal their specific properties. By analyzing the scattering spectra of S&P financial time-series and turbulent jets, we show that none of the mathematical models presented captures all properties of these complex time-series.

Section 6 defines maximum entropy models conditioned by scattering spectra values. We generate time-series according to these models with the microcanonical sampling algorithm in [12]. We show that these generative models can approximate fractional Brownian motions, multifractal random walks and Hawkes processes but also S&P financial time-series or turbu-

lent jets. The code used in numerical experiments is available at [https://github.com/RudyMorel/scattering\\_spectra](https://github.com/RudyMorel/scattering_spectra).

**Notations:** We write  $M^T$  the adjoint and hence the complex conjugate of the transposed matrix  $M$ . We write  $\hat{x}(\omega)$  the Fourier transform of  $x(t)$ .

## 2. Multi-scale Moments

We consider a multi-scale random process  $X(t)$  whose increments are stationary. Gaussian models are maximum entropy models conditioned by first and second order moments. In order to capture non-Gaussian and self-similar properties, one can compute higher order moments of increments. The section reviews the scaling properties of these moments.

### 2.1. Self-Similarity and Power Spectrum

If  $X(t)$  is stationary then its increments are stationary but the reverse is not always true. For example, a Brownian motion  $X$  has stationary increments but  $\mathbb{E}\{X(t)\}$  and  $\mathbb{E}\{X^2(t)\}$  depend on  $t$  [17]. The increments of a random process  $X(t)$  at intervals  $2^j \in \mathbb{R}^+$  for  $j \in \mathbb{R}$  are written

$$\delta_j X(t) = X(t) - X(t - 2^j).$$

The lag  $2^j$  can also be interpreted as a scale parameter. We suppose that  $\delta_j X$  is stationary for any  $j \in \mathbb{R}$ .

Mandelbrot et al. [18] introduced a strong definition of self-similarity from the joint distribution of increments. A process  $X$  is said to be self-similar [18] up to a maximum scale  $2^J$  if for all  $\ell \geq 0$  there exist real random variables  $A_\ell$  which are log infinitely divisible and independent of  $X$  such that

$$\left\{ \delta_j X(t) \right\}_{j \leq J, t \leq N} \stackrel{d}{=} A_\ell \left\{ \delta_{j-\ell} X(2^{-\ell} t) \right\}_{j \leq J, t \leq N}. \quad (1)$$

This equality is in distribution, which means that joint probability distributions of random variables on the left and right hand-sides are equal for any  $(j_1, \dots, j_n)$  and  $(t_1, \dots, t_n)$  with  $n > 0$ . Increments thus have joint distributions which are invariant to dilation, up to random multiplicative factors. The maximum scale  $2^J$  is called the *integrable scale*. It may be fixed, in that case we assume that  $X(t)$  and  $X(t - \tau)$  are nearly independent for  $\tau \gg 2^J$ .

If increments are stationary then their auto-correlation  $\mathbb{E}\{\delta_j X(t) \delta_{j'} X(t - \tau)\}$  only depends on  $\tau$ . By renormalizing its Fourier transform along  $\tau$ , one

can mathematically define a generalized power spectrum  $P_X(\omega)$  of  $X$  [17]. The non-stationarity of  $X$  appears as a singularity of  $P_X(\omega)$ , which tends to  $\infty$  at  $\omega = 0$ . This power spectrum specifies second order moments of increments.

With a scaling argument, one can prove [17] that self-similar processes have a power spectrum which is also self-similar and thus has a power-law scaling

$$P_X(\omega) = c_2 |\omega|^{-\zeta_2-1} \quad (2)$$

which is singular at  $\omega = 0$ .

## 2.2. Increment High Order Moments

First and second order moments define Gaussian maximum entropy models. In order to build non-Gaussian multi-scale models, one can compute  $q$  order moments of increments, if they exist:

$$\forall j \in \mathbb{R} \quad , \quad \mathbb{E}\{|\delta_j X(t)|^q\}. \quad (3)$$

For self-similar processes, these multi-scale moments have a power-law scaling

$$\mathbb{E}\{|\delta_j X(t)|^q\} = \tilde{c}_q 2^{j\zeta_q}. \quad (4)$$

If  $X$  is Gaussian and self-similar then one can verify that  $\zeta_q$  is linear in  $q$  [17]. It results that any non-linear dependency of  $\zeta_q$  as a function of  $q$  implies that  $X$  is not Gaussian. This was initially proposed by Kolmogorov as a test to detect non-Gaussian properties in turbulent flows. Under appropriate hypotheses, the multifractal theory [7] proves that (3) specifies the pointwise Holder regularity of  $X$ , through a spectrum of singularity.

The moment power-law scaling (4) is a weak form of self-similarity which can be tested statistically. On the other hand, the strong self-similarity definition (1) is highly restrictive, often not satisfied, and impossible to be tested on a single realization. Appendix B shows that the strong distribution self-similarity (1) implies the weak moment self-similarity (4). This scaling is simple to test numerically but is a relatively weak characterization of self-similarity. The high order increment moments (4) remain unchanged when computed on  $X(-t)$ , and hence do not detect time-asymmetries. They do not either capture dependencies of increments at different scales  $2^j$ . Section 3 introduces a stronger wide-sense definition of self-similarity, which relies on multi-scale moments that depend upon joint time-scale dependencies of  $X$ .

### 2.3. Estimation and Wavelet Transform

Defining consistent estimators of moments with fast convergence is necessary to compute maximum entropy models from a single realization of  $X$ . It has been proved that a wavelet transform yields nearly optimal estimators of second order moments for self-similar processes [19, 20, 21, 22]. We thus replace increments by a wavelet transform, whose properties are briefly reviewed.

#### 2.3.1. Wavelet transform

A wavelet  $\psi(t)$  has a fast decay away from  $t = 0$ , polynomial or exponential for example, and a zero-average  $\int \psi(t) dt = 0$ . We normalize  $\int |\psi(t)|^2 dt = 1$ . The wavelet transform computes the variations of a signal  $x$  at each scale  $2^j$  with

$$Wx(t, j) = x \star \psi_j(t) \quad \text{where} \quad \psi_j(t) = 2^{-j} \psi(2^{-j}t).$$

If  $\psi = \delta(t) - \delta(t-1)$  then it computes signal increments  $WX(t, j) = \delta_j X(t)$ . To relate regularity properties of signals from their wavelet coefficients, it is necessary to use wavelets having a better frequency localization than a difference of Diracs [7]. We use a complex wavelet  $\psi$  having a Fourier transform  $\hat{\psi}(\omega) = \int \psi(t) e^{-i\omega t} dt$  which is real, and whose energy is mostly concentrated at frequencies  $\omega \in [\pi, 2\pi]$ . It results that  $\hat{\psi}_j(\omega) = \hat{\psi}(2^j \omega)$  is non-negligible mostly in  $\omega \in [2^{-j}\pi, 2^{-j+1}\pi]$ . We suppose that  $\psi$  has  $m \geq 1$  vanishing moments, which means that  $|\hat{\psi}(\omega)| = O(|\omega|^m)$  in the neighborhood of  $\omega = 0$ . We will refer to the modulus and complex phase of  $Wx(t, j)$  as the *amplitude* and *phase* of the complex wavelet coefficient.

In the following we shall restrict the scales  $2^j$  to dyadic scales, and hence  $j$  to integers. The wavelet transform  $W$  satisfies an energy conservation law [23] specified in Appendix A, which implies that it is invertible.

All numerical calculations below are performed with a complex Battle-Lemarié wavelet [24, 25], restricted to positive frequencies. Figure 1 shows the real and imaginary parts of  $\psi$  as well as its Fourier transform. It has an exponential decay away from  $t = 0$ , it has  $m = 4$  vanishing moments and satisfies an energy conservation law (A.1). If the input signal is sampled at  $t \in \mathbb{Z}$  then we can only compute wavelet coefficients for  $2^j > 1$  and hence  $j \geq 1$ .

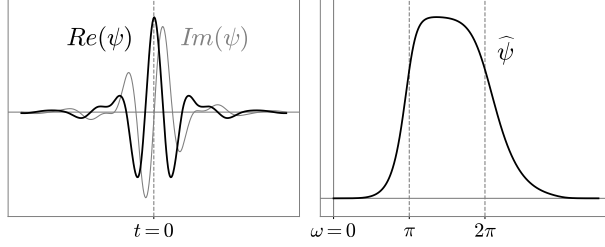


Figure 1: Left: complex Battle-Lemarié wavelet  $\psi(t)$  as a function of  $t$ . Right: Fourier transform  $\hat{\psi}(\omega)$  as a function of  $\omega$ .

### 2.3.2. Wavelet covariance and spectrum

Since  $\int \psi_j(t) dt = 0$  it results that  $\mathbb{E}\{X \star \psi_j(t)\} = 0$ . If  $X$  has stationary increments then one can show [17] that wavelet coefficients are jointly stationary. Their covariance across time and scale can be written from the power spectrum of  $X$ :

$$\mathbb{E}\{X \star \psi_j(t) X \star \psi_{j-a}(t - 2^j \tau)^*\} = \frac{1}{2\pi} \int P_X(\omega) \hat{\psi}(2^j \omega) \hat{\psi}^*(2^{j-a} \omega) e^{i\tau 2^j \omega} d\omega, \quad (5)$$

for time-lag  $2^j \tau \in \mathbb{R}$  and scale-lag  $a \in \mathbb{Z}$ . This covariance becomes negligible when  $|a| > 0$  for which the supports of  $\hat{\psi}(\omega)$  and  $\hat{\psi}(2^a \omega)$  barely overlap. Indeed, the phases of  $X \star \psi_j$  and  $X \star \psi_{j-a}$  vary at different rates, which cancels their correlation. For processes  $X$  with a power-law decaying power-spectrum (2), one can prove that such covariance has a polynomial decay away from  $\tau = 0$  and an exponential decay away from  $a = 0$  [20]. As shown on figure 4a, these coefficients are negligible for distant scales  $|a| > 1$  and have small non-zero values for  $|a| = 1$  because wavelets have a small frequency overlap.

The diagonal covariance values define a *wavelet spectrum*:

$$\sigma_W^2(j) = \mathbb{E}\{|X \star \psi_j(t)|^2\} = \frac{1}{2\pi} \int P_X(\omega) |\hat{\psi}(2^j \omega)|^2 d\omega. \quad (6)$$

It integrates  $P_X(\omega)$  over the frequency intervals  $[2^{-j}\pi, 2^{-j+1}\pi]$ , where  $\hat{\psi}(2^j \omega)$  is mostly supported. It does not depend upon  $t$  because of stationarity, and is thus estimated by an empirical average

$$\tilde{\sigma}_W^2 = \text{Ave}_t (|X \star \psi_j(t)|^2). \quad (7)$$

### 2.3.3. Self-similar wavelet coefficients

If  $X$  is strongly self-similar according to (1) then Appendix B derives that

$$\forall \ell \geq 0, \left\{ X \star \psi_j(t) \right\}_{j \leq J, t \leq N} \stackrel{d}{=} A_\ell \left\{ X \star \psi_{j-\ell}(2^{-\ell}t) \right\}_{j \leq J, t \leq N}. \quad (8)$$

Appendix B also proves that wavelet moments of self-similar processes have the same scaling properties as increments in (4). For all  $q$  such that the moments are defined, there exists  $c_q$  such that

$$\forall j, \quad \mathbb{E}\{|X \star \psi_j|^q\} = c_q 2^{j\zeta_q}. \quad (9)$$

## 3. Dependencies Across Scales with Phase-Modulus Wavelet Correlations

We saw that wavelet coefficients of stationary processes are nearly uncorrelated across scales. Yet, next section shows that non-Gaussian processes have strong dependencies across scales. Section 3.2 captures these dependencies by correlating complex wavelet coefficients and their modulus. Section 3.3 shows that self-similar processes have a normalized phase-modulus wavelet correlation matrix which is invariant to scale shift. This invariance defines a wide-sense self-similarity.

### 3.1. Scale Dependencies as a Trace of Non-Gaussianity

Section 2.3 showed that if  $X$  has stationary increments then  $X \star \psi_j(t)$  and  $X \star \psi_{j'}(t')$  are nearly uncorrelated if  $|j - j'| > 1$ . If  $X$  is Gaussian then these coefficients are jointly Gaussian so it implies that they are independent. On the contrary, we will now see that non-Gaussian time-series exhibits crucial dependency across scales. This dependency provides important information on non-Gaussian properties of  $X$ .

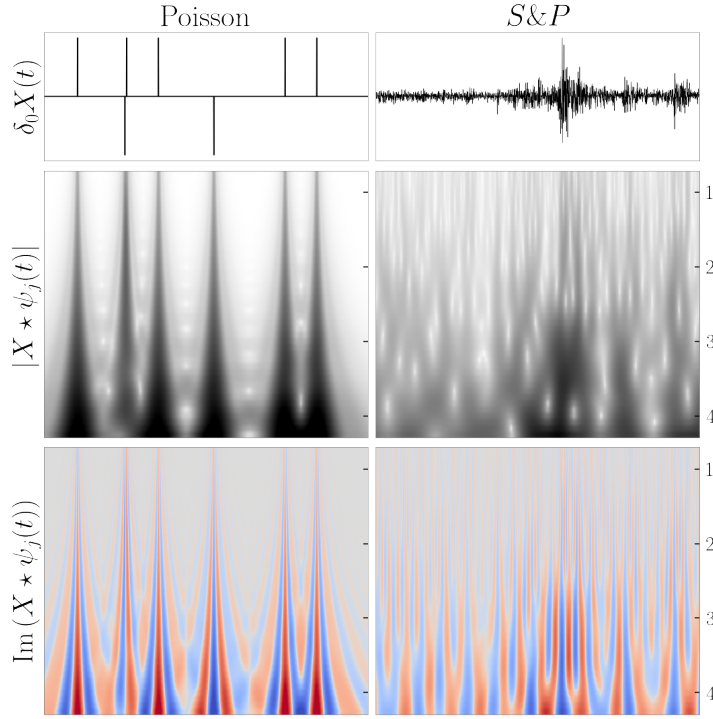
Figure 2 shows the wavelet transform of S&P financial signal, and of a Poisson process whose increments have a random sign. They are calculated with the complex Battle-Lemarié wavelet. Diracs and bursts of activity in the financial signal create high amplitude wavelet coefficients, which propagate across scales. It induces strong dependencies between wavelet modulus across scales. These dependencies also appear in the wavelet transform phase. Diracs produce high amplitude wavelet coefficients whose phase propagates regularly across octaves, when  $j$  increases by 1 or more. On the contrary,

streaks show that the different scales are correlated

vertical streaks, means a burst at one scale tend to appear across several scales

Phase pattern of the wavelet coefficients - red= positive, blue = negative

Repeating patterns across adjacent scales mean phase correlations



Daily returns

j refers to the scale, small j = fine, and large j = coarse structure

Dark bands = large values strong local fluctuations

in gaussian processes scales are independence hence you would not see any streaks

Figure 2: Top: increments  $\delta_0 X(t)$  of a signed Poisson process and the financial S&P daily increments from 03/01/2000 to 10/10/2018. Middle: wavelet modulus  $|X \star \psi_j(t)|$ . The vertical axis corresponds to the log-scale index  $j$  which is real. Dark color represents large values. These modulus have dependencies across scales produced by Diracs or bursts of activity. Bottom: imaginary part of  $X \star \psi_j(t)$ . Red and blue colors represent negative and positive values respectively. It shows that localized structures such as Dirac create correlation of phases across octaves, when  $j$  increases by 1 or more.

financial bursts of activity have a phase that is randomly modified from one octave to the next. Correlations when  $j$  increases by less than 1 are due to correlations between the wavelets themselves.

To understand this dependency phenomenon, consider a localized pattern  $f(t)$  in the neighborhood of  $t = 0$ , such as a Dirac. It is randomly translated to define  $X(t) = f(t - U)$ , where  $U$  is a random variable uniformly distributed in  $[0, 1]$ . Its wavelet coefficients  $X \star \psi_j(t) = f \star \psi_j(t - U)$  are centered at  $t = U$  at all scales  $2^j$ , and are thus highly dependent. Their amplitude and phase are a signature of the translated pattern  $f$ . It illustrates the importance of wavelet coefficient dependencies across scales, for non-Gaussian processes.

### 3.2. Joint Phase-Modulus Correlations Across Scales

This section introduces a joint wavelet phase-modulus correlation matrix which captures dependencies of wavelet coefficients across scales. Complex wavelet coefficients have a negligible correlation at different scales because they are supported in different frequency bands. Their correlation is thus canceled by phase fluctuations which occur at different rates. We first realign their frequency support with a modulus, and then compute their correlation. Correlations of wavelet coefficient modulus were first studied by Portilla and Simoncelli [16]. Their properties are analysed in [14, 26]. The joint phase-modulus correlation matrix also preserves phase information, by correlating wavelet coefficients with and without phases. It partly characterizes phase alignments across scales.

#### 3.2.1. Non-zero correlations by removing complex phase

Eliminating the phase with a modulus can introduce correlations across scales. Indeed, let us remind that the cross-spectrum  $P_{Y,Z}(\omega)$  of two jointly stationary random processes  $Y(t)$  and  $Z(t)$  is the Fourier transform of their cross-correlation

$$P_{Y,Z}(\omega) = \int \mathbb{E}\{Y(t)Z(t-\tau)^*\} e^{-i\tau\omega} d\tau,$$

and the Cauchy-Schwarz inequality proves that

$$|P_{Y,Z}(\omega)|^2 \leq P_Y(\omega) P_Z(\omega). \quad (10)$$

The cross-correlation of  $Y(t)$  and  $Z(t-\tau)$  is therefore zero if their power spectra do not overlap. Applied to  $Y = X \star \psi_j$  and  $Z = X \star \psi_{j-a}$ , it verifies once again that they are essentially uncorrelated if  $a \neq 0$ . Indeed, their power spectrum do not overlap, as illustrated in Figure 3. However, we now show that the power spectrum of  $Y = X \star \psi_j$  and  $Z = |X \star \psi_{j-a}|$  or of  $Y = |X \star \psi_j|$  and  $Z = |X \star \psi_{j-a}|$  can overlap. They can thus have non-zero correlations, after suppressing their mean.

The power spectrum  $P_X(\omega)|\widehat{\psi}(2^{j-a}\omega)|^2$  of  $X \star \psi_{j-a}$  has a support mostly concentrated in  $[2^{-j+a}\pi, 2^{-j+a+1}\pi]$ . A modulus eliminates the phase of  $X \star \psi_{j-a}$  which oscillates at the center frequency  $3 \times 2^{-j+a-1}\pi$ . As a consequence, the power spectrum of  $|X \star \psi_{j-a}|$  is centered at  $\omega = 0$ , and its energy is mostly concentrated in  $[-2^{-j+a}\pi, 2^{-j+a}\pi]$  [14, 26]. This is shown by Figure 3. It results that the spectrum of  $X \star \psi_j$  and  $|X \star \psi_{j-a}|$  do overlap if  $a > 0$ , and the spectrum of  $|X \star \psi_j|$  and  $|X \star \psi_{j-a}|$  overlap for any  $a$  since they are both centered at  $\omega = 0$ .

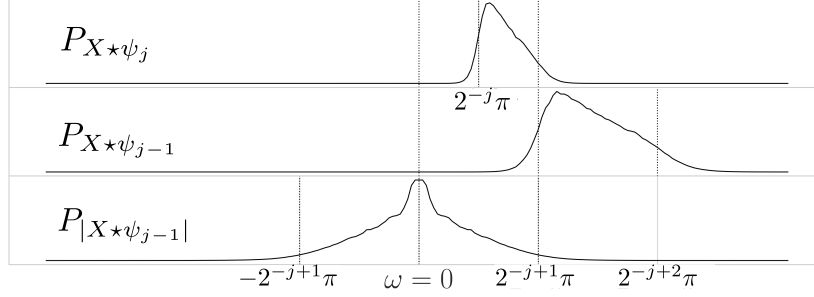


Figure 3: Top: power spectrum of  $X \star \psi_j$  for the S&P time-series. It is mostly concentrated in  $[2^{-j}\pi, 2^{-j+1}\pi]$ . Middle: power spectrum of  $X \star \psi_{j-1}$ . Bottom: power spectrum of  $|X \star \psi_{j-1}|$  is mostly concentrated in  $[-2^{-j+1}\pi, 2^{-j+1}\pi]$  and strongly overlaps with the power spectrum of  $X \star \psi_j$ .

### 3.2.2. Joint phase-modulus correlation matrix

Let us write  $\rho(z) = (z, |z|)$  for any  $z \in \mathbb{C}$ . We now show how to represent the dependencies of wavelet coefficients from joint phase-modulus correlation matrix of

$$\rho W X = (W X, |W X|) = \left( X \star \psi_j(t), |X \star \psi_j(t)| \right)_{t,j}.$$

If  $X$  is sampled at intervals normalized to 1 and is of size  $N$  then we compute wavelet coefficients over  $\log_2 N$  scales  $1 < 2^j \leq N$  corresponding to  $1 \leq j \leq \log_2 N$ . If  $X$  is stationary then  $\mathbb{E}\{\rho W X\} = (\mathbb{E}\{W X\}, \mathbb{E}\{|W X|\})$  does not depend upon  $t$ . Without loss of generality we suppose that  $\mathbb{E}\{X(t)\} = 0$  so  $\mathbb{E}\{W X(t)\} = 0$ .

The coefficients of the joint phase-modulus correlation matrix  $\mathbb{E}\{\rho W X (\rho W X)^*\}$  are

$$\left( \mathbb{E}\{\rho W x(t, j) \rho W x(t - 2^j \tau, j - a)^*\} \right)_{t, \tau, j, a}.$$

They do not depend upon  $t$  because  $X$  is stationary. We estimate them from a single realization of  $X$  with a time average

$$\text{Ave}_{t < N} (\rho W x(t, j) \rho W x(t - 2^j \tau, j - a)^*).$$

The correlation matrix  $\mathbb{E}\{\rho W X (\rho W X)^T\}$  is composed of four submatrices

$$\begin{pmatrix} \mathbb{E}\{W X W X^T\} & \mathbb{E}\{W X |W X|^T\} \\ \mathbb{E}\{|W X| W X^T\} & \mathbb{E}\{|W X| |W X|^T\} \end{pmatrix}$$

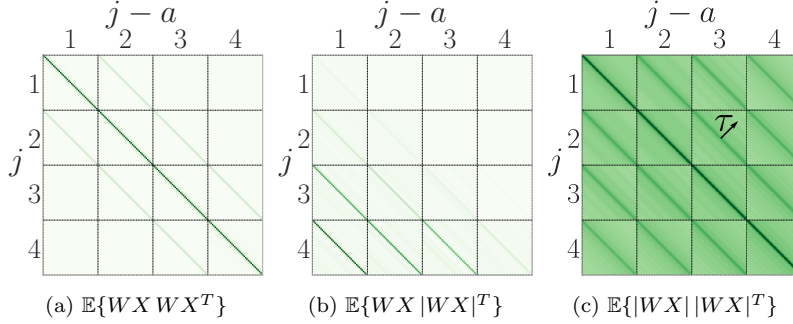


Figure 4: Modulus of the joint phase-modulus correlation  $\mathbb{E}\{\rho WX (\rho WX)^*\}$  for S&P signal. The diagonal of this matrix ( $\tau = 0, a = 0$ ) is the wavelet spectrum  $\sigma_W^2(j)$  which is not constant. To remove such normalization effect, we plot the matrix where each coefficient  $(t, \tau, j, a)$  is divided by  $\sigma_W(j)\sigma_W(j-a)$ . For the 3 matrices, each subblock is a Toeplitz correlation matrix along time  $(t, t-2^j\tau)$ , for scales  $(j, j-a)$  fixed, because of time stationarity. All correlation values are constant when  $j$  varies, which is a mark of wide-sense self-similarity.

We show that  $\mathbb{E}\{WXWX^T\}$  and  $\mathbb{E}\{WX|WX|^T\}$  are sparse matrices. On the other hand, we shall see that  $\mathbb{E}\{|WX||WX|^T\}$  may not be sparse. The coefficients of the wavelet auto-correlation matrix  $\mathbb{E}\{WXWX^T\}$  are

$$\mathbb{E}\{X \star \psi_j(t) X \star \psi_{j-a}(t - 2^j\tau)^*\}. \quad (11)$$

Section 2.3 explains that we can approximate it by its diagonal values which define the wavelet spectrum. For a signal of size  $N$ , since  $1 \leq j \leq \log_2 N$  we only estimate  $\log_2 N$  wavelet spectrum coefficients  $\mathbb{E}\{|X \star \psi_j(t)|^2\}$ .

The off-diagonal matrix  $\mathbb{E}\{WX|WX|^T\}$  is a correlation between complex wavelet coefficients and their modulus, that we shall call *phase-modulus correlation*.

$$\mathbb{E}\{X \star \psi_j(t) |X \star \psi_{j-a}(t - 2^j\tau)|\} \quad (12)$$

Since  $\mathbb{E}\{X \star \psi_j\} = 0$  these correlations are also covariance coefficients. Figure 4b shows that wavelet phase-modulus correlations are non-negligible for a scale shift  $a \geq 0$  and  $\tau = 0$ . This is expected because the power spectrum of  $X \star \psi_j$  and  $|X \star \psi_{j-a}|$  have an overlapping support. When  $\tau \neq 0$  they become negligible because of random phase fluctuations. Since  $1 \leq j-a < j \leq \log_2 N$ , we only estimate  $2^{-1} \log_2^2 N$  wavelet phase-modulus cross-spectrum coefficients  $\mathbb{E}\{X \star \psi_j(t) |X \star \psi_{j-a}(t)|\}$ .

The coefficients of the wavelet *modulus auto-correlation*  $\mathbb{E}\{|WX||WX|^T\}$

are

$$\mathbb{E}\{|X \star \psi_j(t)| |X \star \psi_{j-a}(t - 2^j \tau)|\}. \quad (13)$$

Figure 4c shows that the wavelet modulus correlation is nearly a full matrix. Their covariance is obtained by subtracting the modulus means. These covariances are also a priori non-zero for all scale shift  $a$ , because the power spectra of  $|X \star \psi_j|$  and  $|X \star \psi_{j-a}|$  have overlapping supports. They can be non-negligible for large time shift  $\tau$ , because all phases have been eliminated. The number of time shifts is nearly equal to the signal size  $N$ . For  $1 \leq j - a \leq a \leq \log_2 N$ , there are about  $2^{-1} N \log_2^2 N$  potentially non-negligible wavelet modulus correlations, which is too large to estimate them directly from a single realization of  $X$ . Section 4 shows that one can reduce the number of correlation coefficients to  $\log_2^3 N$ , by applying a second wavelet transform on  $|X \star \psi_j(t)|$  before calculating its auto-correlation.

### 3.2.3. Non-Gaussian properties

Phase and modulus wavelet correlations can be analyzed as particular cases of phase harmonic correlations introduced in [14]. It captures non-Gaussian properties proved in [26]. The following proposition transpose these results in our context. It proves that the existence of non-negligible wavelet modulus covariances across scales is a mark of non-Gaussianity. We write  $\text{Cov}\{A, B\} = \mathbb{E}\{AB^*\} - \mathbb{E}\{A\} \mathbb{E}\{B\}^*$ .

**Proposition 1.** *If  $X$  is Gaussian and  $\widehat{\psi}_j \widehat{\psi}_{j-a} = 0$  then for all  $\tau$*

$$\text{Cov}\{\rho Wx(t, j), \rho Wx(t - 2^j \tau, j - a)\} = 0.$$

We indeed saw that if  $\widehat{\psi}_j \widehat{\psi}_{j-a} = 0$  then  $X \star \psi_j(t)$  and  $X \star \psi_{j-a}(t - 2^j \tau)$  are uncorrelated (5). If  $X$  is Gaussian then they are also Gaussian and hence independent. Applying a modulus preserves this independence and thus produces covariance coefficients which remain zero, proving the second equality. The condition  $\widehat{\psi}_j \widehat{\psi}_{j-a} = 0$  is verified up to a very small error for  $|a| > 1$ . For  $|a| = 1$ , the supports of  $\widehat{\psi}_j$  and  $\widehat{\psi}_{j-a}$  have a small overlap so the product is small but not zero. It follows from the proposition that non-zero covariance coefficients across distant scales evidence that  $X$  is not Gaussian.

Time-asymmetry is another form of non-Gaussianity, often produced by causality phenomena. Let  $R$  be the time reversal operator  $Rx(t) = x(-t)$ . A process  $X$  is said to be time-reversible if the probability distributions of  $RX$  and  $X$  are equal. Gaussian stationary processes are time-reversible.

The following proposition shows that time-reversibility can be detected from phase correlation coefficients.

**Proposition 2.** *If  $X$  is time-reversible then the joint wavelet modulus correlation has a Hermitian symmetry along  $\tau$ :*

$$\mathbb{E}\{\rho W X(t, j) \rho W X(t - 2^j \tau, j - a)^*\} = \mathbb{E}\{\rho W X(t, j)^* \rho W X(t + 2^j \tau, j - a)\}. \quad (14)$$

Time-reversibility means that  $X$  and  $RX$  have the same distribution. Since  $(RX) \star \psi_j(t) = X \star \psi_j(-t)^*$  it implies the equality (14). This Hermitian symmetry is always satisfied by the wavelet auto-correlation coefficients (11) even if  $X$  is not time-reversible, but not necessarily by phase-modulus correlations and modulus auto-correlation coefficients (12,13). If they do not have the Hermitian symmetry (14) then  $X$  is not time-reversible, and hence non-Gaussian.

### 3.3. Wide-Sense Self-Similarity

Self-similarity is defined in (1) and (8) on the process distributions, which are high-dimensional objects, on increments or wavelet coefficients. Such properties cannot be tested statistically on a single realization. Section 2 gives necessary conditions (4,9) over the marginals of increments and wavelet coefficients, which are simple to verify but provides relatively weak characterization of self-similarity. The same difficulty appears to test that a process has a stationary distribution. It cannot be tested statistically on a single realization. Conditions on marginals impose that the probability distributions of  $X(t)$  for each  $t$  does not depend upon  $t$ . It can be tested numerically but it is a weak condition. More powerful characterizations of stationarity impose that  $\mathbb{E}\{X(t)\}$  and the auto-correlation of  $X$  are invariant to time shift. The process  $X$  is then said to be wide-sense stationary. We follow the same approach for self-similarity by imposing a scale shift invariance on a normalized joint phase-modulus correlation matrix.

Self-similarity of increments distributions (1) or wavelet coefficients (8) are defined up to random multiplicative factors  $A_\ell$ . We eliminate these multiplicative factors with a normalized phase-modulus correlation matrix where each correlation coefficient of  $\mathbb{E}\{\rho W X (\rho W X)^*\}$  is normalized by a product of standard deviations given by  $\sigma_W^2 = \mathbb{E}\{|X \star \psi_j(t)|^2\}$ :

$$C_{\rho W}(\tau; j, a) = \frac{\mathbb{E}\{\rho W X(t, j) \rho W X(t - 2^j \tau, j - a)^*\}}{\sigma_W(j) \sigma_W(j - a)}.$$

A wavelet transform introduces explicitly a scaling parameter  $2^j$ . However, the correlations of wavelet coefficients  $WX$  vanish across scales, and thus can not be used directly to identify self-similarity across scales. We define a notion of wide-sense self-similarity as a translation and scale invariance of the mean and correlation matrix of  $(WX, |WX|)$ .

**Theorem 3** (Wide-sense self-similarity). *If  $X$  is self-similar up to the scale  $2^J$  in the sense of (8) then there exist  $c_1, c_2, \zeta_1, \zeta_2$  such that for all  $j \leq J$*

$$\mathbb{E}\{|X \star \psi_j(t)|\} = c_1 2^{j\zeta_1}, \quad (15)$$

$$\mathbb{E}\{|X \star \psi_j(t)|^2\} = c_2 2^{j\zeta_2}. \quad (16)$$

and for all  $\tau, j \leq J, a$ ,

$$C_{\rho W}(\tau; j, a) = C_{\rho W}(\tau; 0, a). \quad (17)$$

The theorem is proved in Appendix C. A process  $X$  which satisfies the properties (15,16,17) is said to be wide-sense self-similar. The appendix proves that self-similarity implies wide-sense similarity. Similarly to moment self-similarity (9), wide-sense self-similarity imposes the existence of scaling exponents  $\zeta_q$ , but only for  $q = 1$  and  $q = 2$ . The scaling exponent  $\zeta_2$  specifies the decay of the wavelet spectrum  $\sigma_W(j) = \mathbb{E}\{|X \star \psi_j(t)|^2\}$ . The ratio between first and second order moments is a sparsity measure on wavelet coefficients

$$s_W(j) = \frac{\mathbb{E}\{|X \star \psi_j(t)|\}}{\sigma_W(j)}. \quad (18)$$

If  $X$  is wide-sense self-similar then  $s_W^2(j) = c_s 2^{j\zeta_s} \leq 1$ , where  $c_s = c_1^2 c_2^{-1}$  and  $\zeta_s = 2\zeta_1 - \zeta_2 \geq 0$ . The lower  $s_W^2(j)$  the sparser  $X \star \psi_j(t)$ . The exponent  $\zeta_s$  is a sparsity rate which governs the increase of sparsity when the scale decreases. If  $X$  is Gaussian then  $\zeta_s = 0$ . If  $\zeta_s > 0$  then the sparsity of  $X \star \psi_j$  increases as  $j$  decreases. The constant  $c_s$  is a sparsity multiplicative factor. If  $X$  is Gaussian then  $X \star \psi_j$  is also Gaussian and one can verify that  $c_s = \pi/4$ , which is the ratio between first and second order moments of complex Gaussian random variables.

Wide-sense self-similarity also imposes that the normalized phase-modulus correlation matrix  $C_{\rho W}$  depends only on time shift  $\tau$  and scale shift  $a$ . This is a powerful second order condition, which is sufficient to reveal existence of important self-similar properties in time-series. It applies to the non-negligible coefficients of each of the three submatrices of  $C_{\rho W}$ . Over diagonal

wavelet auto-correlation coefficients, it is already specified by (16). Over non-negligible phase-modulus cross-spectrum coefficients, it imposes that

$$C_{W|W|}(0; j, a) = \frac{\mathbb{E}\{WX(t, j) |WX(t, j - a)|\}}{\sigma_W(j) \sigma_W(j - a)}$$

does not depend upon  $j$ . These coefficients are estimated by replacing each expected value by an average on  $t$ . For self-similar processes, since these moment do not depend on  $j$ , we can further improve this estimation by averaging them over scales. It defines an *scale invariant phase-modulus cross spectrum*

$$\overline{C}_{W|W|}(a) = \text{Ave}_j C_{W|W|}(0; j, a). \quad (19)$$

For wavelet modulus auto-correlations (13), these conditions are translated into conditions over a scattering cross-spectrum which is introduced in the next section.

Processes such as fractional Brownian motion [1] or multifractal random walk [27] are self-similar and therefore wide-sense self-similar. Self-similarity cannot be tested statistically, whereas wide-sense self-similarity is a correlation property which can be estimated. Figure 4 shows that the S&P financial signal is wide-sense self-similar. Indeed  $\log \mathbb{E}\{|X \star \psi_j|\}$  and  $\log \mathbb{E}\{|X \star \psi_j|^2\}$  have a linear decay along  $j$ , and the normalized correlation  $C_{\rho W}$  is constant along its diagonals when  $j$  varies.

## 4. Scattering Cross-Spectrum

Section 3.2 explains that there are too many wavelet modulus auto-correlation coefficients to estimate them from a single realization of  $X$ . We introduce a low-dimensional approximation of this auto-correlation matrix by applying a second wavelet transform, which defines a scattering transform [28]. The resulting scattering covariance is nearly diagonal and its spectra can be estimated from a single realization.

### 4.1. Diagonal Scattering Cross-Correlation

Section 2.3 explains that if  $X$  has self-similarity properties then its auto-correlation matrix is well approximated by a diagonal matrix after applying a wavelet transform. Similarly, instead of computing directly the auto-correlation of  $|WX| = (|X \star \psi_j(t)|)_{j,t}$  we will compute the auto-correlation of its wavelet transform.

Applying a second wavelet transform  $W$  on  $|WX|$  defines a scattering transform  $SX = W|WX|$ , with

$$SX(t; j, k) := |X \star \psi_j| \star \psi_k(t).$$

It is non-negligible only if  $k > j$ . Indeed, the Fourier transform of  $|X \star \psi_j|$  is mostly concentrated in  $[-2^{-j}\pi, 2^{-j}\pi]$ . If  $k \leq j$  then it does not intersect the frequency interval  $[2^{-k}\pi, 2^{-k+1}\pi]$  where the energy of  $\widehat{\psi}_k$  is mostly concentrated, in which case  $SX(t; j, k) \approx 0$ .

Since  $SX = W|WX|$ , its auto-correlation is

$$\mathbb{E}\{SX SX^T\} = W \mathbb{E}\{|WX| |WX|^T\} W^T, \quad (20)$$

which specifies  $\mathbb{E}\{|WX| |WX|\}$  because  $W$  is invertible. The coefficients of  $\mathbb{E}\{SX SX^T\}$  are

$$\mathbb{E}\{|X \star \psi_j| \star \psi_k(t) |X \star \psi_{j-a}| \star \psi_{k'}(t - \tau)^*\}$$

for all time  $t$ , time shift  $\tau$ , first wavelet scale  $j$  and scale shift  $a$ , and second wavelet scales  $k, k'$ . We impose that  $k > j$  and  $k' > j - a$  otherwise the scattering coefficients are negligible. Applying (10) to  $Y(t) = |X \star \psi_j| \star \psi_k(t)$  and  $Z = |X \star \psi_{j-a}| \star \psi_{k'}(t - \tau)$  shows that this correlation is negligible if  $k \neq k'$  because the spectra of  $Y$  and  $Z$  barely overlap. Indeed  $\widehat{\psi}_k$  and  $\widehat{\psi}_{k'}$  are concentrated over non-overlapping frequency intervals. Correlations for  $k = k'$  have a fast polynomial decay away from  $\tau = 0$  [20] when the cross-spectrum of the modulus is regular, and we thus shall only consider these scattering correlations for  $\tau = 0$ . Scattering correlations are thus calculated only for  $k = k' = j - b$  with  $b < 0$  and  $\tau = 0$ .

We incorporate the normalization of the wavelet modulus auto-correlation by the wavelet spectrum  $\mathbb{E}\{|X \star \psi_j|^2\}$  to the scattering correlations. The normalized diagonal scattering coefficients for  $k = k' = j - b$  define a *scattering cross-spectrum* whose coefficients are:

$$C_S(j, a, b) = \frac{\mathbb{E}\{|X \star \psi_j| \star \psi_{j-b}(t) |X \star \psi_{j-a}| \star \psi_{j-b}(t)^*\}}{\sigma_W(j) \sigma_W(j - a)}.$$

Since  $\psi_{j-b}$  has a Fourier transform mostly supported in  $[2^{-j+b}\pi, 2^{-j+b+1}\pi]$ ,  $C_S(j, a, b)$  can be interpreted as the cross-spectrum of  $|X \star \psi_j|$  and  $|X \star \psi_{j-a}|$  integrated over this frequency interval. These cross-spectra specify intermittency phenomena which appear when the wavelet modulus correlations (13)

remain large on a long-range of time. If the modulus are uncorrelated in time across scales then  $C_S(j, a, b)$  is nearly constant along  $b$  for  $j, a$  fixed. On the contrary, if  $C_S(j, a, b)$  has a fast decay in  $b$  then it implies that the modulus have long range correlations in time.

If  $X$  is of size  $N$  then there are at most  $\log_2 N$  scales indices  $j, a$  and  $b$ . it shows that the scattering transform can provide an approximation of  $N \log_2^2 N$  wavelet modulus auto-correlation coefficients with  $\log_2^3 N$  scattering cross-spectrum coefficients.

#### 4.2. Properties

The following proposition derives from Proposition 2 that the imaginary part of  $C_S$  captures time-asymmetry properties of  $X$ . The proof is in Appendix D.

**Proposition 4.** *If  $X$  is Gaussian and  $\widehat{\psi}_j \widehat{\psi}_{j-a} = 0$  then*

$$C_S(j, a, b) = 0.$$

*If  $X$  is time-reversible then for all  $j, a$  and  $b$  the imaginary part satisfies*

$$\text{Im } C_S(j, a, b) = 0.$$

The following theorem proves that the self-similarity condition (17) on  $C_{\rho W}$  can be evaluated on non-zero scattering coefficients.

**Theorem 5.** *The scale invariance (17) of  $C_{\rho W}$  implies that*

$$\forall j \leq J, \quad C_S(j, a, b) = C_S(0, a, b). \quad (21)$$

The theorem is proved in Appendix D. This condition on scattering cross-spectrum coefficients is necessary and almost sufficient to guarantee the scale invariance of the normalized wavelet modulus auto-correlation  $\mathbb{E}\{|WX| |WX|^T\}$ . To do so we would also need to impose an invariance condition on the off-diagonal coefficients of  $C_S$  but these coefficients have mostly a negligible amplitude. In the following, we shall thus systematically replace  $\mathbb{E}\{|WX| |WX|^T\}$  by the scattering cross-spectrum correlation  $C_S$  and assess the scale invariance from (21). They are estimated by replacing expected values by a time averaging. For self-similar processes,  $C_s$  does not depend on  $j$ . These invariant coefficients are thus estimated by averaging them over scales. It defines a *scale invariant scattering cross-spectrum*

$$\overline{C}_S(a, b) = \text{Ave}_j C_S(j, a, b). \quad (22)$$

For a signal of size  $N$ , there are at most  $\log_2^2 N$  such coefficients.

## 5. Numerical Dashboard for multi-scale Processes

We show that the scattering spectra, defined as the multi-scale moments (6,18,19,22), specify intermittency, time-asymmetries and self-similar properties of multi-scale processes. Next section studies standard mathematical models of multi-scale processes, and the following section considers numerical financial and turbulence time-series.

### 5.1. Models of self-similar processes

We consider Brownian motions, Poisson processes, multifractal random walks and Hawkes processes. To analyze their self-similarity properties we display and analyze their scattering spectra, composed of

- $\sigma_W^2(j)$ : wavelet spectrum (6) shown in Figure 5a,
- $s_W^2(j)$ : wavelet sparsity factor (18) in Figure 5b,
- $\overline{C}_{W|W|}(a)$ : scale invariant phase-modulus cross-spectrum (19) in Figure 6.
- $\overline{C}_S(a, b)$ : scale invariant scattering cross-spectrum (22) in Figure 7.

Table 1 gives the power-law decay parameters of  $\sigma_W^2$  and  $s_W^2$ .

Expected values are estimated with empirical averages over  $N$  samples in time, as in (7). If  $X$  has a finite integrable scale  $s$ , which means that  $X(t)$  and  $X(t')$  are independent if  $|t - t'| > s$  (see section 2.1), then we set the maximum wavelet scale  $2^J$  to be equal to  $s$ . When the signal size  $N$  goes to  $\infty$ , time average estimators are consistent estimators of expected values. Indeed, if the wavelet has a support of size  $\alpha$  then all scattering spectra coefficients are expected values of operators whose support sizes are at most  $2\alpha 2^J$ . One can thus verify that each empirical estimator averages at least  $N(2\alpha s)^{-1}$  blocks of independent coefficients, which have the same mean because  $X$  is stationary. These empirical estimators thus converge to the expected value when  $N$  increases, with a variance which decays at least like  $2\alpha s N^{-1}$ .

In this section, we consider multi-scale processes whose integrable scale is not necessarily finite. The maximum wavelet scale  $2^J$  is chosen to be smaller than the signal size  $N$  in order for large scale coefficients to be well estimated. Time average estimators of scattering spectra coefficients can then provide consistent estimators at all the scales smaller than  $2^J$ . The

figure

Sparsity measures how bursty/spiky the signal is at each scale  
- flat curve means same level of burst

Falling curve means bursts get strong or weaker depending on scale

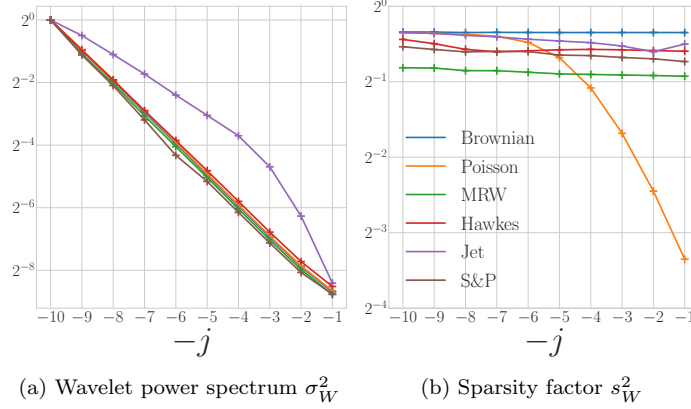


Figure 5: (a) Power spectrum  $\sigma_W^2(j)$  in (6) as a function of  $-j$ , which is a log frequency index. (b) Sparsity factor  $s_W^2(j)$  in (18) as a function of  $-j$ . Each process is shown with a different color. Non-linear curves reveal absence of self-similarity for Poisson process and the turbulent jet at fine scales.

variance decay of our estimators is not guaranteed mathematically, but it is verified numerically.

	Brown.	Poisson	MRW	Hawkes	Jet	S&P
$\zeta_2$	1.0	1.0	1.0	1.0	0.66	1.0
$c_s$	0.79	$\times$	0.66	0.65	0.67	0.61
$\zeta_s$	0.00	$\times$	0.016	0.013	0.025	0.016

Table 1: Self-similarity parameters  $\zeta_2, c_s, \zeta_s$ , for different multi-scale processes. For the jet,  $\zeta_2, c_s, \zeta_s$  are given on the self-similar range.

*Fractional Brownian motion.* It is a Gaussian self-similar process, studied by Mandelbrot [1]. It has stationary increments and a generalized power spectrum  $P_X(\omega) = c_2 |\omega|^{-\zeta_2-1}$ . Computations are performed with  $\zeta_2 = 1$ , which corresponds to a standard Brownian motion. Wavelet sparsity coefficients have a power law decay with  $c_s = \pi/4$  and  $\zeta_s = 0$  in Table 1, because it is a Gaussian process. Propositions 1 and 4 prove that  $\overline{C}_{W|W|}(a) \approx 0$  and  $\overline{C}_S(a, b) \approx 0$  for  $a > 0$ , which is verified in Figure 6 and Figure 7. For  $a = 0$ , we also observe in Figure 7 that  $\overline{C}_S(0, b)$  is constant, which shows that  $|X \star \psi_j(t)|$  are uncorrelated at time increments which are sufficiently large relatively to the scale  $2^j$ . As proved by propositions 2 and 4, the phases of  $C_{W|W|}$  and  $C_S$  are zero in Figures 6 and 7, because a Gaussian process is time-reversible. Since Brownian motions are self-similar, the phase-modulus

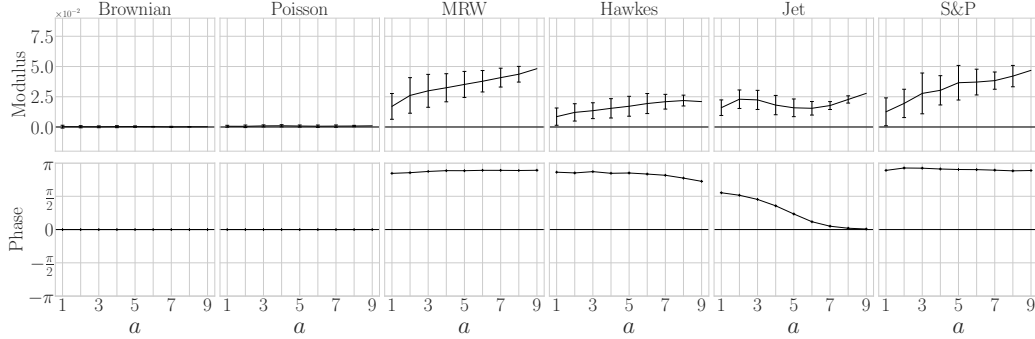


Figure 6: Modulus and phase of the scale invariant phase-modulus cross-spectrum  $\bar{C}_{W|W|}(a)$  (19). A skewed MRW and a quadratic Hawkes are shown. Error bars represent the mean-square variations of  $C_{W|W|}(0; j, a)$  around  $\bar{C}_{W|W|}(a)$ . Non-zero phase coefficients reveal time-asymmetry.

cross-spectrum  $C_{W|W|}$  and the scattering cross-spectrum  $C_S$  remain constant across scales  $2^j$ .

*Poisson process.* It is a jump process having independent stationary increments. The number of jumps in an interval is proportional to the intensity  $\lambda$ . We further suppose that each jump is positive or negative with a probability  $1/2$ . Its power spectrum has a power law decay with  $\zeta_2 = 1$ , but it is non-Gaussian and not self similar, which clearly appears in its scattering spectra. Figure 5b shows that  $\log s_W^2(j)$  in (24) has a slope which varies as a function of  $-j$ . Indeed, if  $2^j \ll \lambda^{-1}$  then [29] proves that  $s_W^2(j) \sim 2^j$  because

$$\mathbb{E}\{|X \star \psi_j(t)|\}^2 \sim \lambda^2 2^{2j} \quad \text{and} \quad \mathbb{E}\{|X \star \psi_j(t)|^2\} \sim \lambda^2 2^j,$$

whereas if  $2^j \gg \lambda^{-1}$  then  $s_W^2(j) \sim 1$  because  $X \star \psi_j(t)$  converges in distribution to a Gaussian white noise of variance  $\lambda 2^j$  as  $2^j$  goes to  $\infty$  [29], and hence

$$\mathbb{E}\{|X \star \psi_j(t)|\}^2 \sim \lambda 2^j \quad \text{and} \quad \mathbb{E}\{|X \star \psi_j(t)|^2\} \sim \lambda 2^j.$$

The non-self-similarity of Poisson process is also revealed by the large variations of  $C_S(j, a, b)$  along  $j$ , which appears as large error bars in Figure 7.

*Multifractal Random Walk (MRW).* It is a non-Gaussian self-similar process, whose increments have scaling exponents  $\zeta_q$  in (4) that are quadratic in  $q$ .

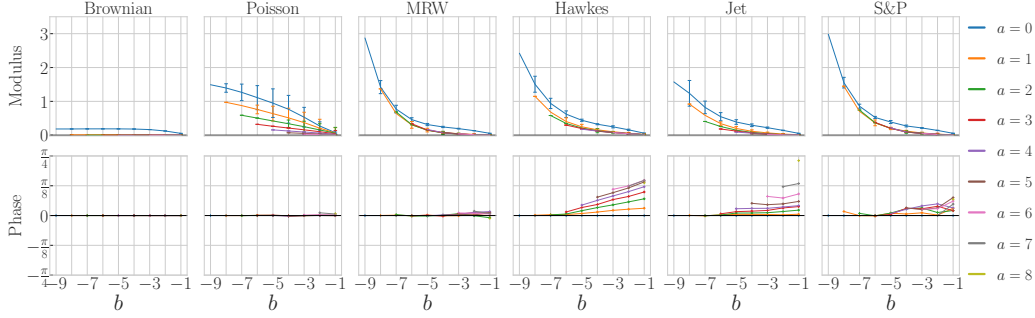


Figure 7: Modulus and phase of the scale invariant scattering cross-spectrum  $\overline{C}_S(a, b)$  (22) as a function of  $b$ . A skewed MRW and a quadratic Hawkes are shown. The parameter  $b$  is a log-frequency. Each color curve corresponds to a different scale shift  $a$ . Error bars represent the mean-square variations of  $C_S(j, a, b)$  around  $\overline{C}_S(a, b)$ . Non-zero phases reveal time-asymmetry of wavelet modulus.

Increments are computed as a product of the increment  $\delta_j B$  of a Brownian motion with a log-normal process

$$\delta_j X(t) = \delta_j B(t) e^{\Omega_j(t)}.$$

The Gaussian process  $\Omega_j(t)$  has an auto-correlation with a slow logarithmic decay:

$$\text{Cov}\{\Omega_j(2^j t), \Omega_j(2^j t')\} = \lambda^2 \ln \theta(|t - t'|),$$

where  $\theta$  decreases linearly and is specified in [27]. Since  $\Omega_j$  is highly correlated in time, it creates wavelet modulus that are also highly correlated in time with bursts of activity. The parameter  $\lambda$  governs the intensity of this intermittency. If  $\lambda = 0$  then the multifractal random walk is a Brownian motion. For MRW one can prove [27] that  $\zeta_s = \lambda^2$ , so Table 1 recovers the value  $\lambda^2 = \zeta_s = 0.016$ .

The scale invariant scattering cross-spectrum  $\overline{C}_S(a, b)$  is the power spectrum of the modulus auto-correlation at two scales shifted by  $a$ , where  $b$  is a log-frequency index. Long range correlation of wavelet modulus appears in Figure 7, which shows that  $\overline{C}_S(a, b)$  has a power-law decay for each  $a$ .

A skewed multifractal random walk incorporates a time-asymmetry by imposing that increments in the past are correlated with the future volatility [30], in order to reflect the so-called leverage effect [31, 32]. This volatility is defined in finance as the mean square average of increments over a fixed period of time. It amounts to replacing the Gaussian process  $\Omega_j$  by  $\Omega_j - k \star$

$\delta_j B$  where  $k(t)$  is a strictly causal power-law convolution kernel. The faster  $K$  decreases the shorter the scale of asymmetry. Figure 6 shows  $\overline{C}_{W|W|}(a)$  for skewed MRW. As expected from Proposition 2, this time-asymmetry is revealed by the non-zero phase of  $\overline{C}_{W|W|}$ , which implies that its imaginary part is also non-zero.

*Hawkes process.* It is a non-homogeneous, causal self-excited point process [33, 34], where each jump is positive or negative with probability 1/2. The jump intensity  $\lambda_t$  depends on the average of the past-jumps with power-law decaying kernel  $h$

$$\lambda_t = \lambda_\infty + h \star |dN|_t,$$

where  $dN_s$  is the signed jump measure and  $h \star |dN|_t := \int_{-\infty}^t h(t-s)|dN_s|$ . A linear feedback term  $l \star dN_t$  can be added to account for correlation between past signed increments and future volatility.

A quadratic Hawkes model [35] introduces time-asymmetric modulus dependencies with a causal quadratic feedback of kernel  $k(t)$

$$\lambda_t = \lambda_\infty + h \star |dN|_t + l \star dN_t + |k \star dN_t|^2.$$

If  $\int |h(t)| dt < 1$  then a quadratic Hawkes is stationary, and if  $\int (|h(t)| + |k(t)|^2) dt < 1$  the mean intensity  $\bar{\lambda}$  is finite [36]. In numerical simulation, as in [35] we set  $h(t) \sim |t|^{-1.2}$ ,  $g(t) \sim e^{-0.01|t|}$  and  $k(t) \sim e^{-0.03|t|}$  with  $\int (|h(t)| + |k(t)|^2) dt = 0.9$ . As expected from Proposition 4, Figure 7 shows that for this quadratic Hawkes,  $\overline{C}_S(a, b)$  has a non-zero phase which reveals its modulus time-asymmetry [37].

## 5.2. Analysis of multi-scale Time-Series

Brownian motions, multifractal random walks and Hawkes processes are used as models of multi-scale time-series, particularly in finance and turbulence [1, 18, 38, 39, 33]. The next two paragraphs analyze the scattering spectra of financial and turbulent time-series. Expected values are computed with time average estimators, which introduce an estimation error.

### 5.2.1. Finance

Finding stochastic models of financial time-series is important to compute the price of financial instruments which mitigate financial risks, such as options. A Brownian motion is a simple first order model, on which the Black and Scholes option pricing formula is based. However, many studies have

shown strong deviations to Gaussianity, including the existence of bursts of activity and crises. Multifractal random walks, Hawkes processes and rough volatility models are among the most popular models used to capture non-Gaussian properties of financial markets [40, 38, 41, 42, 34, 35]. We consider the American stock index S&P log-prices sampled over 5 minutes from January 2000 to October 2018. A standard preprocessing is performed and is described in Appendix E. The resulting signal has  $N = 7.5 \cdot 10^5$  samples.

Figure 5 and Table 1 show that S&P has a wavelet spectrum decay exponent  $\zeta_2 = 1$ , which is the same as a Brownian motion. However, its sparsity factor  $c_s = 0.61 \neq \pi/4 \approx 0.79$  and exponent  $\zeta_s = 0.016 \neq 0$ , which shows that this time-series is clearly not Gaussian. These values are matched by the MRW model, and by a Hawkes process with calibrated parameters.

The S&P scale invariant phase-modulus cross-spectrum  $\overline{C}_{W|W|}$  in Figure 6 is similar to a skewed multifractal random walk (SMRW), with a strong time-asymmetry shown by a non-zero phase, related to the leverage effect [31, 32]. The amplitude of  $|\overline{C}_S(a, b)|$  in Figure 7 is similar for the S&P and the MRW. However, the phase of  $\overline{C}_S(a, b)$  is non-zero for the S&P, which proves that wavelet modulus are also asymmetric in time. This is not well captured by the SMRW model. Such an effect is referred to as the Zumbach effect in finance, which can be explained from causal agent reactions to past trends [37, 41, 35]. Such a modulus time-asymmetry appears in the quadratic Hawkes model, but  $|\overline{C}_S(a, b)|$  has different variations along the  $a$  direction for S&P and for the Hawkes model, presumably because of our choice of an exponentially decaying kernel  $k(t)$  (which was calibrated on intraday data only). This analysis of the scattering spectra shows that none of these mathematical models fully capture all statistical properties of the S&P time-series.

Financial signals are often believed to be self-similar. We can estimate whether the S&P satisfies the wide-sense self-similarity properties of Theorem 3. The wavelet spectrum and sparsity factors in Figure 5 do indeed have a power law decay. Wide-sense self-similarity also imposes that  $C_{W|W|}(0; j, a)$  and  $C_S(j, a, b)$  do not depend upon the scale  $j$ . Figures 6 and 7 display their mean-square variations along  $j$  as error bars. They are of the same order as the estimation variance of each coefficients. We thus conclude that S&P time-series has no significant deviation from wide-sense self-similarity.

### 5.2.2. Turbulence

Kolmogorov introduced in 1941 a self-similar Gaussian model of turbulence [43, 44, 45], with an asymptotic analysis of Navier-Stokes equations at

high Reynolds numbers. This analysis predicts that the projection of the velocity field on a line is a stationary process whose power spectrum has a power-law decay with  $\zeta_2 = 2/3$ . However, turbulence time-series are highly non-Gaussian, and Kolmogorov's initial theory was then refined to take into account intermittency and the presence of vortices [46, 47].

We study the scattering spectra of experimental velocity measurements along a single direction, measured over a turbulent gaseous helium jet at low temperature, with a high Reynolds number equal to 929 [48]. This time-series has  $N = 3.5 \cdot 10^7$  samples and is thus much larger than the S&P time-series, providing more accurate estimators of the scattering spectra. The non-zero phase of  $\overline{C}_{W|W|}(a)$  and  $\overline{C}_S(a, b)$  in Figures 6 and 7 shows a time-asymmetry, which results from the directionality of the jet propulsion. The quadratic Hawkes provides the best model of  $\overline{C}_S(a, b)$  but it fails to accurately represent  $\overline{C}_{W|W|}(a)$ .

It thus appears that none of the existing mathematical model provides accurate models of this turbulent time-series.

Turbulence time-series may be self-similar on a frequency range limited by the Reynolds number at low frequencies and by the fluid viscosity at high frequencies. The wavelet power spectrum and sparsity factors in Figure 5 are indeed self-similar up to the finest scale  $j = 2$ , which is the diffusion scale created by the fluid viscosity. The self-similarity error bars in Figure 6 and Figure 7 are of the same order as for the S&P. However, their amplitude is statistically significant in this case because this time-series is 50 times larger than the S&P time-series and the estimation variance is thus much smaller. It shows that this turbulent time-series has significant deviations from wide-sense self-similarity.

## 6. Maximum Entropy scattering spectra Models

Brownian motions, multifractal random walks and Hawkes models are defined by a fixed number of parameters which does not depend upon their size  $N$ . They can be calibrated on data, but we saw they are typically too restrictive to capture all important properties of multi-scale time-series. On the other hand, neural network models [49, 50] have a considerable flexibility but they can not be trained from a single time-series because the number of parameters is much larger than  $N$ .

This section introduces maximum entropy models computed from the scattering spectra energy vector  $\Phi(x)$  of dimension smaller than  $\log_2^3 N$  and

hence much smaller than  $N$ . This energy vector is specified in the next section. It provides consistent estimators of the scattering spectra for random processes having a finite integral scale. We study the approximation of multi-scale time-series from such maximum entropy models. Section 6.3 shows that scattering spectra models are sufficiently flexible to approximate a wide range of mathematical processes, as well as real financial and turbulence data.

### 6.1. Scattering Spectra Energy Vector

We define maximum entropy models of the form  $p_\theta(x) = Z_\theta^{-1} e^{-\theta^T \Phi(x)}$  for a certain  $\theta \in \mathbb{R}^M$ , where the energy vector  $\Phi(x)$  computes the scattering spectra estimation for any time-series  $x$  of dimension  $N$ .

The scattering spectra vector  $\Phi(x)$  is normalized by wavelet spectrum coefficients, which are constants estimated from a realization  $\tilde{x}$  of the random process  $X$  that we want to model

$$\tilde{\sigma}_W^2(j) = \text{Ave}_{t < N} (|\tilde{x} \star \psi_j(t)|^2).$$

If  $X(t)$  has a finite integrable scale (see section 2.1), the estimator  $\tilde{\sigma}_W$  of  $\sigma_W$  is consistent as  $N$  goes to  $\infty$ .

Low-frequencies are captured by the low-pass filter  $\varphi_J = \psi_{J+1}$  defined in (A.2). The scattering spectra energy is defined by four families of coefficients

$$\Phi(x) = (\Phi_1(x), \Phi_2(x), \Phi_3(x), \Phi_4(x)).$$

The first family provides  $J$  order 1 moment estimators squared, corresponding to wavelet sparsity coefficients (18)

$$\Phi_1(x) = \frac{(\text{Ave}_{t < N} |x \star \psi_j(t)|)^2}{\tilde{\sigma}_W^2(j)}. \quad (23)$$

The  $J + 1$  normalized second order wavelet spectrum associated to  $x$  are computed by

$$\Phi_2(x) = \frac{\text{Ave}_{t < N} |x \star \psi_j(t)|^2}{\tilde{\sigma}_W^2(j)}. \quad (24)$$

There are  $J(J + 1)/2$  wavelet phase-modulus cross-spectrum coefficients

$$\Phi_3(x) = \frac{\text{Ave}_{t < N} (x \star \psi_j(t) |x \star \psi_{j-a}(t)|)}{\tilde{\sigma}_W(j) \tilde{\sigma}_W(j - a)}. \quad (25)$$

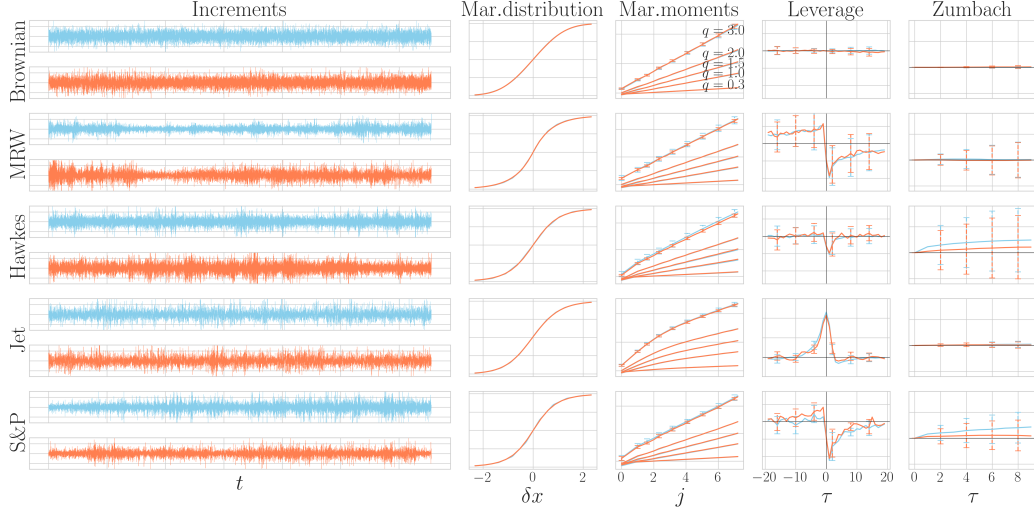


Figure 8: Synthesis results. (Left) Increments  $\delta_0 X(t)$  of original (blue) and generated (orange) time-series. (Right) Using the same color code: cumulative marginal distribution function  $F(\delta x)$ , log of marginal moments  $\text{Ave}_{t < N}(|\delta_j X(t)|^q)$  as a function of  $j$ , leverage  $\mathcal{L}_j(\tau)$  and Zumbach integral, for the Brownian, MRW (skewed), Hawkes (quadratic), turbulent jet and S&P. Leverage and Zumbach are shown for a specific  $j$ . Error bars show the standard deviation for test moments of order greater than 3. Though it is based on order 1 and order 2 moments, our model is able to capture higher order moments used to reveal scaling properties as well as time-asymmetries.

Finally, it includes less than  $J(J+1)(J+2)/6$  scattering cross-spectrum

$$\Phi_4(x) = \frac{\text{Ave}_{t < N}(|x \star \psi_j| \star \psi_{j-b}(t) |x \star \psi_{j-a}| \star \psi_{j-b}^*(t))}{\tilde{\sigma}_W(j) \tilde{\sigma}_W(j-a)}. \quad (26)$$

The dimension of  $\Phi(x)$  is smaller than  $J^3 \leq \log_2^3 N$  as soon as  $N \geq 8$ , this is much smaller than the dimension  $N$  of  $x$ .

### 6.2. Model Validation with Test Moments

This section evaluates numerically the precision of maximum entropy models defined from the scattering spectra energy  $\Phi(x)$ , to approximate the probability distributions of real mathematical processes and real data. The maximum entropy model  $p_\theta(x) = Z_\theta^{-1} e^{-\theta^T \Phi(x)}$  is sampled with a standard microcanonical algorithmic approach reviewed in Appendix F. It is computed with a gradient descent which avoids estimating the macrocanonical

parameters  $\theta$ . The choice of the maximum wavelet scale  $2^J$  amounts to defining an integrable scale equal to  $2^J$  and hence nearly independent coefficients at distances larger than  $2^J$ . Indeed, the scattering spectra model then does not impose any constraint on coefficients whose distance are much larger than  $2^J$  so the entropy maximisation yields a random process whose coefficients are nearly independent at such distances.

Assessing the precision of a model from a dataset of samples drawn from an unknown distribution is an ill-defined problem. A maximum entropy model constrains the values of moments  $\mathbb{E}\{\Phi(X)\}$ . One may find errors by comparing moments which are not explicitly constrained. Such *test moments* are estimated on the original time-series and on time-series generated by the model itself. Following [51], we describe test moments used in the finance literature to identify important differences relatively to Brownian motions. Figure 8 compares moment values obtained from a scattering spectra model with the original time-series. The following test moments are computed on increments  $\delta_j X(t) = X(t) - X(t - 2^j)$  which are stationary. While our model considers dyadic scales  $j \in \mathbb{Z}$ , we consider all scales  $2^j \in \mathbb{R}^+$  for test moments.

*Cumulative distribution*  $F(\delta x)$  of increments  $\delta_0 X(t) = X(t) - X(t - 1)$  at the finest scale.

*Marginal high order moments* described in (3) are estimated by time average  $\text{Ave}_{t < N} (|\delta_j X(t)|^q)$  at all scales  $2^j$ , for  $0.3 \leq q \leq 3$ . The multi-scale properties of these moments is reviewed in Section 2.2.

*Leverage moments* measure asymmetric dependencies between past and future increments in finance [31, 32]. A leverage correlation for a time shift  $\tau$  is an order 3 moment, at any scale  $2^j$ :

$$\mathcal{L}_j(\tau) = \text{Ave}_{t < N} (\delta_j X(t - \tau) |\delta_j X(t)|^2).$$

If  $X$  has a time-reversible distribution then  $\mathcal{L}_j(\tau) = -\mathcal{L}_j(-\tau)$ . Results are shown in Figure 8 at an intermediate scale  $2^j = 159$  that corresponds to the day in the case of S&P. The same scale is taken for all processes except for the Jet for which we take  $2^j = 1$ .

*Zumbach moments* evaluate the time-asymmetry of the volatility in finance [37, 41]. The volatility is the energy of increments over a time period of size  $2^j$

$$\Sigma_j^2(t) = \text{Ave}_{t-2^j \leq u < t} |\delta_0 X(u)|^2.$$

A Zumbach moment for a time shift  $\tau$  is an order 4 moment at a scale  $2^j$

$$\mathcal{Z}_j(\tau) = \text{Ave}_{t < N} (|\delta_j X(t - \tau)|^2 \Sigma_j^2(t)).$$

If  $X$  is time-reversible then  $\mathcal{Z}_j(-\tau) = \mathcal{Z}_j(\tau)$ . To evaluate the time-asymmetry, Figure 8 shows  $\int_0^t (\mathcal{Z}_j(s) - \mathcal{Z}_j(-s)) ds$  as a function of  $t$  for a scale  $2^j = 159$ , that corresponds to the day for S&P. This asymmetry coefficient on an order 4 moment is typically estimated with a large variance as we shall see.

### 6.3. Generation from Scattering Spectra Models

Figure 8 gives results of scattering spectra models computed from a single realization of a Brownian motion, a skewed MRW, a quadratic Hawkes process, a turbulence jet and the S&P financial signal. It displays realizations generated by these models and compares test moments. For financial and turbulence data, the syntheses recover signals of size  $N = 7.10^5$  with models computed over  $J = 11$  scales. The resulting scattering spectra model has 375 parameters. Figure 8 shows that all test moments of order 2 or below are perfectly reproduced by scattering spectra models, for Brownian motion, MRW and Hawkes as well as for the turbulence jet and S&P financial data. Marginal moments of order  $1/3 \leq q \leq 3$  are captured by our model on all processes, which is in accordance with the well reproduced cdf.

For test moments of order 3 or 4, including the leverage and Zumbach effects that capture time-asymmetries, we represent the variance of estimators with an error bar, which is quite large for the Zumbach effect. Leverage is well captured for MRW, Hawkes, Jet, and remains within the estimation error bar for S&P. Zumbach integral estimations have a much larger variance. The main information is in the sign of this integral, when significant. This test moment is again reproduced on both Hawkes and S&P within the estimation error. Its high variance clearly shows the importance of using low order moments, even for time-asymmetries. scattering spectra models reveal such non-Gaussian properties with a modulus and moments of order 1 and 2. In the case of the S&P, we believe that any remaining discrepancy for the Zumbach effect comes from our somewhat naive treatment of closing periods during the night.

## 7. Conclusion

We introduced the scattering spectra which give an interpretable low-dimensional representation of processes having stationary increments. It

gaussian, means that the values produces at each time step is drawn from some normal distribution

captures their power spectrum, multi-scale sparsity, and the dependencies of wavelet coefficients phase and modulus across scales. Wide-sense self-similar signals have scattering spectra which are invariant to scale shifts, and thus define a representation of even lower dimension.

For a time series of size  $N$ , this scattering spectra is at most of dimension  $\log_2^3 N$ . We showed numerically that it reveals potential non-Gaussianity and self-similar properties. This was demonstrated on mathematical multi-scale models such as fractional Brownian motions, multifractal random walks and Hawkes processes, but also on real time-series in finance and turbulence. Maximum entropy scattering spectra models capture essential multi-scale dependency properties and can be efficiently sampled with a microcanonical approach.

Scattering spectra models are related to generative convolutional neural networks based on covariance matrices [52]. Similarly to a one-hidden layer convolutional neural network, it computes a cascade of two convolutions and a pointwise non-linearity. The network filters are wavelets which are not learned. It provides a much lower dimensional representation of random processes than usual deep convolutional neural networks, and it is furthermore interpretable. However, it only applies to signals which are stationary or have stationary increments.

## Appendix A. Wavelet Transform Properties

We impose that the wavelet  $\psi$  satisfies the following energy conservation law called Littlewood-Paley equality

$$\forall \omega > 0, \quad \sum_{j=-\infty}^{+\infty} |\widehat{\psi}(2^j \omega)|^2 = 1. \quad (\text{A.1})$$

A Battle-Lemarié wavelet [24, 25] is an example of such wavelet. The wavelet transform is computed up to a largest scale  $2^J$  which is smaller than the signal size  $N$ . The signal lower frequencies in  $[-2^{-J}\pi, 2^{-J}\pi]$  are captured by a low-pass filter  $\varphi_J(t)$  whose Fourier transform is

$$\widehat{\varphi}_J(\omega) = \left( \sum_{j=J+1}^{+\infty} |\widehat{\psi}(2^j \omega)|^2 \right)^{1/2}. \quad (\text{A.2})$$

One can verify that it has a unit integral  $\int \varphi_J(t) dt = 1$ . To simplify notations, we write this low-pass filter as a last scale wavelet  $\psi_{J+1} = \varphi_J$ , and  $Wx(t, J+1) = x \star \psi_{J+1}(t)$ . By applying the Parseval formula, we derive from (A.1) that for all  $x$  with  $\|x\|^2 = \int |x(t)|^2 dt < \infty$

$$\|Wx\|^2 = \sum_{j=-\infty}^{J+1} \|x \star \psi_j\|^2 = \|x\|^2.$$

The wavelet transform  $W$  preserves the norm and is therefore invertible, with a stable inverse.

Properties of signal increments are carried over to wavelet coefficients by observing that wavelet coefficients are obtained by filtering signal increments  $\delta_j X(t) = X(t) - X(t - 2^j)$  with a dilated integrable filter:

$$X \star \psi_j(t) = \delta_j X \star \theta_j(t) \quad \text{where} \quad \theta_j(t) = 2^{-j} \theta(2^{-j} t), \quad (\text{A.3})$$

where filter  $\theta$  is obtained from  $\psi$  through  $\widehat{\theta}(\omega) = \widehat{\psi}(\omega) / (1 - e^{-i\omega})$ . This is because  $1 - e^{i2^j \omega}$  is the Fourier transform of  $\delta(t) - \delta(t - 2^j)$ , the filter that creates increments. From (A.3) we get that if  $\delta_j X(t)$  is stationary then  $X \star \psi_j(t)$  is also stationary.

## Appendix B. Proof that strong distribution self-similarity implies weak moment self-similarity.

Let  $X$  be a stationary process that is self-similar according to (1). For  $q \in \mathbb{R}$ , marginal moments are written  $S(q, j) = \mathbb{E}\{|\delta_j X(t)|^q\}$ . They do not depend upon  $t$ . For all  $\ell \geq 0, j \leq J$ , self-similarity implies that marginal distributions are equal:  $\delta_j X(2^\ell t) \stackrel{d}{=} A_\ell \delta_{j-\ell} X(t)$ . On order  $q$  moments, since  $A_\ell$  is independent from  $X$  this yields

$$S(q, j) = \mathbb{E}\{A_\ell^q\} S(q, j - \ell).$$

Since the factors  $(A_\ell)_\ell$  are log-infinitely divisible, for all  $\ell_1, \ell_2 \geq 0$   $\mathbb{E}\{A_{\ell_1+\ell_2}^q\} = \mathbb{E}\{A_{\ell_1}^q\} \mathbb{E}\{A_{\ell_2}^q\}$ . This implies that  $\log \mathbb{E}\{A_\ell^q\}$  is linear in  $\ell$  which means there exists  $\zeta_q$  such that  $\mathbb{E}\{A_\ell^q\} = 2^{j\zeta_q}$ . By defining  $\tilde{S}(q, j) = 2^{-j\zeta_q} S(q, j)$ , we obtain  $\tilde{S}(q, j) = \tilde{S}(q, j - \ell)$  for all  $j \leq J, \ell \geq 0$ , which implies that  $\tilde{S}(q, j)$  is equal to a constant  $\tilde{c}_q$  which does not depend on  $j$

$$S(q, j) = \mathbb{E}\{|\delta_j X(t)|^q\} = \tilde{c}_q 2^{j\zeta_q}. \quad (\text{B.1})$$

Let us now establish the same property for wavelet coefficients. According to the self-similarity property (1), wavelet coefficients satisfy (8). Indeed, one has:

$$\begin{aligned} X \star \psi_j(2^\ell t) &= \delta_j X \star \theta_j(2^\ell t) \text{ thanks to (A.3)} \\ &= \delta_j X(2^\ell \cdot) \star \theta_{j-\ell}(t) \text{ as } \theta_j \text{ are dilated filters} \\ &\stackrel{d}{=} A_\ell \delta_{j-\ell} X \star \theta_{j-\ell}(t) \text{ by self-similarity (1)} \\ &= A_\ell X \star \psi_{j-\ell}(t) \text{ thanks to (A.3)} \end{aligned}$$

Increments  $\delta_j X(t)$  are a special case where  $\psi_j = \delta(t) - \delta(t - 2^j)$ . For general wavelets  $\psi_j$  the same proof than for increments holds, there exists  $c_q$  such that for the same  $\zeta_q$ :

$$\mathbb{E}\{|X \star \psi_j(t)|^q\} = c_q 2^{j\zeta_q}.$$

## Appendix C. Proof that strong distribution self-similarity implies wide-sense self-similarity

Let  $X$  be a process with stationary increments that is self-similar and thus satisfies (8). For  $q = 1$  and  $q = 2$ , Appendix B proves that  $\mathbb{E}\{|X \star \psi_j(t)|\} = c_1 2^{j\zeta_1}$  and

$$\sigma_W^2(j) = \mathbb{E}\{|X \star \psi_j(t)|^2\} = c_2 2^{j\zeta_2}, \quad (\text{C.1})$$

which proves (15) and (16).

The equality in distribution (8) implies that for  $\tau, j, a$  fixed we have

$$\begin{aligned} & \left( X \star \psi_j(t), X \star \psi_{j-a}(t - 2^j \tau) \right) \\ & \stackrel{d}{=} A_\ell \left( X \star \psi_{j-\ell}(2^{-\ell} t), X \star \psi_{j-a}(2^{-\ell} t - 2^{j-\ell} \tau) \right) \end{aligned}$$

Applying  $\rho$  and taking expected value gives

$$\begin{aligned} & \mathbb{E}\{\rho W X(t, j) \rho W X(t - 2^j \tau, j - a)\} \\ & = 2^{\ell \zeta_2} \mathbb{E}\{\rho W X(t, j - \ell) \rho W X(t - 2^{j-\ell} \tau, j - a - \ell)\}, \end{aligned}$$

because of stationarity and  $\mathbb{E}\{A_\ell^2\} = 2^{\ell \zeta_2}$ . Normalized correlations  $C_{\rho W}(\tau; j, a)$  are obtained by dividing by  $\sigma_W(j)\sigma_W(j - a)$ . It results from (C.1) that

$$2^{\ell \zeta_2} \sigma_W(j)^{-1} \sigma_W(j - a)^{-1} = \sigma_W(j - \ell)^{-1} \sigma_W(j - a - \ell)^{-1}$$

and hence

$$C_{\rho W}(\tau; j, a) = C_{\rho W}(\tau, j - \ell, a).$$

Taking  $j = \ell$  proves (17).

## Appendix D. Proof of proposition 4 and theorem 5

Let  $X$  be a Gaussian process with stationary increments and assume that  $\widehat{\psi_j} \widehat{\psi_{j-a}} = 0$ . Then for any  $\tau$ ,  $X \star \psi_j(t)$  and  $X \star \psi_{j-a}(t - \tau)$  are decorrelated because their power spectra do not overlap. Since  $X$  is Gaussian these are also Gaussian. It implies that the processes  $X \star \psi_j(t)$  and  $X \star \psi_{j-a}(t)$  are independent. In particular,  $|X \star \psi_j| \star \psi_{j-b}(t)$  and  $|X \star \psi_{j-a}| \star \psi_{j-b}(t)$  are independent. Their correlation is thus zero and  $C_S(j, a, b) = 0$ .

Let  $X$  be a time-reversible process with stationary increments. Thanks to proposition 2 we know that  $C_{\rho W}$  has the Hermitian symmetry:  $C_{\rho W}(\tau; j, a) = C_{\rho W}(-\tau; j, a)^*$ . Let us write  $C_{|W|}$  the subblock of this matrix composed of the normalized modulus auto-correlation

$$C_{|W|}(\tau; j, a) = \frac{\mathbb{E}\{|X \star \psi_j(t)| |X \star \psi_{j-a}(t - 2^j \tau)|\}}{\sigma_W(j)\sigma_W(j - a)}$$

We derive from (20) that the scattering cross-spectrum  $C_S$  satisfies:

$$C_S(j, a, b) = \int_{\tau} C_{|W|}(\cdot; j, a) \star \psi_{-b}(\tau) \psi_{-b}(\tau)^* d\tau. \quad (\text{D.1})$$

With  $Rx(t) = x(-t)$ , the Hermitian symmetry of  $C_{|W|}$  implies that

$$\begin{aligned}
C_S(j, a, b) &= \int_{\tau} RC_{|W|}(\cdot; j, a)^* \star \psi_{-b}(\tau) \psi_{-b}(\tau)^* d\tau \\
&= \int_{\tau} C_{|W|}(\cdot; j, a)^* \star R\psi_{-b}(-\tau) \psi_{-b}(\tau)^* d\tau \\
&= \int_{\tau} C_{|W|}(\cdot; j, a)^* \star \psi_{-b}^*(-\tau) \psi_{-b}(\tau)^* d\tau \\
&= \int_{\tau} C_{|W|}(\cdot; j, a)^* \star \psi_{-b}^*(-\tau) \psi_{-b}(-\tau) d\tau \\
&= C_S^*(j, a, b)
\end{aligned}$$

because  $R\psi_{-b} = \psi_{-b}^*$ . It proves that  $\text{Im } C_S(j, a, b) = 0$  which proves proposition 4.

Under scale invariance (17)  $C_{\rho W}(\tau; j, a) = C_{\rho W}(\tau; 0, a)$ . As a subblock, one has also  $C_{|W|}(\tau; j, a) = C_{|W|}(\tau; 0, a)$ . In particular, the equation (D.1) that expresses  $C_S$  from  $C_{|W|}$  implies  $C_S(j, a, b) = C_S(0, a, b)$  which proves theorem 5.

## Appendix E. Financial data preprocessing

We use a standard preprocessing on S&P data which accounts for missing values, overnight period, intraday seasonality and tick effect. It is performed on 5min increments of S&P from January 3rd 2000 to October 10th 2018 which represents 751 116 values.

Missing values, 17 956 5min increments, are replaced by independent Gaussian values with zero mean and standard deviation observed at this time of the day.

Intraday seasonality is the fact that the volatility is larger at certain typical hours of the day: it is a non-stationary effect. It is removed by dividing 5min increments by the average volatility profile over all days.

The overnight period corresponds to the first bin at the beginning of each day. The corresponding increments are generally larger than other 5min increments. That again creates a non-stationarity effect that can be attenuated by dividing each overnight increment by their average volatility on all days.

Prices of S&P are present on a grid with certain tick size. Hence, 5min increments are discrete with many values equal to 0 or to plus/minus the tick size. This tick effect is present only for high-frequency increments and

hence breaks the scale invariance property at high frequency. To remove it we apply a low-pass filter to  $X$  that amounts to a moving average on small windows of 15 minutes.

## Appendix F. Microcanonical set and sampling

We briefly review a maximum entropy microcanonical approach to sample a maximum entropy model defined from a single realization  $\tilde{x}$  of  $X$  [12]. A microcanonical set of width  $\epsilon$  takes into account the error between  $\Phi(\tilde{x})$  and  $\mathbb{E}\{\Phi(X)\}$

$$\Omega_\epsilon = \{x \in \mathbb{R}^N \mid \|\Phi(x) - \Phi(\tilde{x})\|_2 \leq \epsilon\}.$$

The width  $\epsilon$  is adjusted so that each realization of  $X$  of size  $N$  is in  $\Omega_\epsilon$ , with high probability (in practice, we choose  $\epsilon = 10^{-3}\|\Phi(\tilde{x})\|_2$ ). The probability distribution of  $X$  is thus essentially supported in  $\Omega_\epsilon$ . The width  $\epsilon$  goes to zero when  $N$  increases if  $\Phi(X)$  is a mean-square consistent estimator of  $\mathbb{E}\{\Phi(X)\}$  when  $N$  goes to infinity. The convergence of microcanonical models towards macrocanonical models is a large deviation problem studied in statistical physics in the limit of  $N$  going to infinity [53].

A microcanonical model is the maximum entropy distribution supported in  $\Omega_\epsilon$ . Since  $\Omega_\epsilon$  is bounded, the microcanonical model has a uniform distribution over  $\Omega_\epsilon$ . Sampling the microcanonical model thus amounts to select  $x$  in  $\Omega_\epsilon$ , with a uniform probability. Such a sampling can be calculated by a gradient descent studied in [12]. It progressively transports a Gaussian white noise, which has a maximum entropy, into a distribution supported in  $\Omega_\epsilon$ , with a push-forward map calculated with a gradient descent on  $\ell(x) = \|\Phi(x) - \Phi(\tilde{x})\|_2^2$ .

The initial point  $x_0$  is a realization of a Gaussian white noise. At each iteration  $n$ , it updates  $x_n$  with a gradient descent step

$$x_{n+1} = x_n - \eta \nabla \ell(x_n).$$

The gradient descent is implemented with the L-BFGS-B algorithm. It is proved in [12] that it converges to a measure which approximates the maximum entropy measure. It has the same invariance to time shift as  $\Phi(x)$  similarly to the maximum entropy measure.

## References

- [1] B. B. Mandelbrot and J. W. Van Ness, “Fractional brownian motions, fractional noises and applications,” *SIAM review*, vol. 10, no. 4, pp. 422–437, 1968.
- [2] B. B. Mandelbrot, “Multifractals and  $1/f$  noise wild self-affinity in physics (1963–1976).”
- [3] —, *The fractal geometry of nature*. W. H. Freeman and Co., 1982.
- [4] E. Bacry, J.-F. Muzy, and A. Arneodo, “Singularity spectrum of fractal signals from wavelet analysis: Exact results,” *Journal of statistical physics*, vol. 70, no. 3, pp. 635–674, 1993.
- [5] J.-F. Muzy, E. Bacry, and A. Arneodo, “The multifractal formalism revisited with wavelets,” *International Journal of Bifurcation and Chaos*, vol. 4, no. 02, pp. 245–302, 1994.
- [6] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, “Wavelets for the analysis, estimation, and synthesis of scaling data,” *Self-Similar Network Traffic and Performance Evaluation*, pp. 39–88, 2000.
- [7] S. Jaffard, “Wavelet techniques in multifractal analysis,” PARIS UNIV (FRANCE), Tech. Rep., 2004.
- [8] S. Jaffard, B. Lashermes, and P. Abry, “Wavelet leaders in multifractal analysis,” in *Wavelet analysis and applications*. Springer, 2006, pp. 201–246.
- [9] H. Wendt, S. G. Roux, S. Jaffard, and P. Abry, “Wavelet leaders and bootstrap for multifractal analysis of images,” *Signal Processing*, vol. 89, no. 6, pp. 1100–1114, 2009.
- [10] R. Leonarduzzi, P. Abry, H. Wendt, S. Jaffard, and H. Touchette, “A generalized multifractal formalism for the estimation of nonconcave multifractal spectra,” *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 110–119, 2018.

- [11] U. Frisch and G. Parisi, “Fully developed turbulence and intermittency,” *Proceedings of the International Summer School on Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, p. 84–88, 1985.
- [12] J. Bruna and S. Mallat, “Multiscale sparse microcanonical models,” *Mathematical Statistics and Learning*, vol. 1, no. 3, pp. 257–315, 2019.
- [13] D. R. Brillinger, “An introduction to polyspectra,” *The Annals of mathematical statistics*, pp. 1351–1374, 1965.
- [14] S. Mallat, S. Zhang, and G. Rochette, “Phase harmonic correlations and convolutional neural networks,” *Information and Inference: A Journal of the IMA*, vol. 9, no. 3, pp. 721–747, 2020.
- [15] J. Bruna and S. Mallat, “Invariant scattering convolution networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.
- [16] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *International journal of computer vision*, vol. 40, no. 1, pp. 49–70, 2000.
- [17] V. Pipiras and M. S. Taqqu, *Long-range dependence and self-similarity*. Cambridge university press, 2017, vol. 45.
- [18] B. B. Mandelbrot, A. J. Fisher, and L. E. Calvet, “A multifractal model of asset returns,” 1997.
- [19] P. Flandrin, “Wavelet analysis and synthesis of fractional brownian motion,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 910–917, 1992.
- [20] G. Wornell, “Wavelet-based representations for the  $1/f$  family of fractal processes,” *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1428–1450, 1993.
- [21] E. Masry, “The wavelet transform of stochastic processes with stationary increments and its application to fractional brownian motion,” *IEEE Transactions on Information Theory*, vol. 39, no. 1, pp. 260–264, 1993.

- [22] E. McCoy and A. Walden, “Wavelet analysis and synthesis of stationary long-memory processes,” *Journal of computational and Graphical statistics*, vol. 5, no. 1, pp. 26–56, 1996.
- [23] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.
- [24] G. Battle, “A block spin construction of ondelettes. part i: Lemarié functions,” *Communications in Mathematical Physics*, vol. 110, no. 4, pp. 601–615, 1987.
- [25] P.-G. Lemarié, “Ondelettes à localisation exponentielle,” *J. Math. Pures Appl.*, vol. 67, pp. 227–236, 1988.
- [26] S. Zhang and S. Mallat, “Maximum entropy models from phase harmonic covariances,” *Applied and Computational Harmonic Analysis*, vol. 53, pp. 199–230, 2021.
- [27] E. Bacry, J. Delour, and J. F. Muzy, “Multifractal random walk,” *Phys. Rev. E*, vol. 64, p. 026103, Jul 2001. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevE.64.026103>
- [28] S. Mallat, “Group invariant scattering,” *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, 2012.
- [29] J. Bruna, S. Mallat, E. Bacry, J.-F. Muzy *et al.*, “Intermittent process analysis with scattering moments,” *Annals of Statistics*, vol. 43, no. 1, pp. 323–351, 2015.
- [30] B. Pochart and J.-P. Bouchaud, “The skewed multifractal random walk with applications to option smiles,” *Quantitative finance*, vol. 2, no. 4, p. 303, 2002.
- [31] G. Bekaert and G. Wu, “Asymmetric volatility and risk in equity markets,” *The review of financial studies*, vol. 13, no. 1, pp. 1–42, 2000.
- [32] J.-P. Bouchaud, A. Matacz, and M. Potters, “Leverage effect in financial markets: The retarded volatility model,” *Physical review letters*, vol. 87, no. 22, p. 228701, 2001.
- [33] E. Bacry and J.-F. Muzy, “Hawkes model for price and trades high-frequency dynamics,” *Quantitative Finance*, vol. 14, no. 7, pp. 1147–1166, 2014.

- [34] E. Bacry, I. Mastromatteo, and J.-F. Muzy, “Hawkes processes in finance,” *Market Microstructure and Liquidity*, vol. 1, no. 01, p. 1550005, 2015.
- [35] P. Blanc, J. Donier, and J.-P. Bouchaud, “Quadratic hawkes processes for financial prices,” *Quantitative Finance*, vol. 17, no. 2, pp. 171–188, 2017.
- [36] C. Aubrun, M. Benzaquen, and J.-P. Bouchaud, “On hawkes processes with infinite mean intensity,” *arXiv!*, p. 2112.14161, 2021.
- [37] G. Zumbach, “Time reversal invariance in finance,” *Quantitative Finance*, vol. 9, no. 5, pp. 505–515, 2009.
- [38] E. Bacry, J. Delour, and J.-F. Muzy, “Modelling financial time series using multifractal random walks,” *Physica A: statistical mechanics and its applications*, vol. 299, no. 1-2, pp. 84–92, 2001.
- [39] N. Mordant, J. Delour, E. Léveque, A. Arnéodo, and J.-F. Pinton, “Long time correlations in lagrangian dynamics: a key to intermittency in turbulence,” *Physical review letters*, vol. 89, no. 25, p. 254502, 2002.
- [40] B. B. Mandelbrot, “The variation of certain speculative prices,” *The Journal of Business*, vol. 36, no. 4, p. 394–419, 1963.
- [41] R. Chicheportiche and J.-P. Bouchaud, “The fine-structure of volatility feedback i: Multi-scale self-reflexivity,” *Physica A: Statistical Mechanics and its Applications*, vol. 410, pp. 174–195, 2014.
- [42] J. Gatheral, T. Jaisson, and M. Rosenbaum, “Volatility is rough,” *Quantitative finance*, vol. 18, no. 6, pp. 933–949, 2018.
- [43] A. N. Kolmogorov, “Dissipation of energy in the locally isotropic turbulence,” in *Dokl. Akad. Nauk SSSR A*, vol. 32, 1941, pp. 16–18.
- [44] —, “On degeneration (decay) of isotropic turbulence in an incompressible viscous liquid,” in *Dokl. Akad. Nauk SSSR*, vol. 31, 1 941, pp. 538–540.
- [45] —, “The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers,” *Cr Acad. Sci. URSS*, vol. 30, pp. 301–305, 1941.

- [46] —, “A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high reynolds number,” *Journal of Fluid Mechanics*, vol. 13, no. 1, p. 82–85, 1962.
- [47] U. Frisch, “From global scaling, a la kolmogorov, to local multifractal scaling in fully developed turbulence,” *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, vol. 434, no. 1890, pp. 89–99, 1991.
- [48] O. Chanal, B. Chabaud, B. Castaing, and B. Hébral, “Intermittency in a turbulent low temperature gaseous helium jet,” *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 17, no. 2, pp. 309–317, 2000.
- [49] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “Wavenet: A generative model for raw audio,” *arXiv preprint arXiv:1609.03499*, 2016.
- [50] F. Eckerli, “Generative adversarial networks in finance: an overview,” *Available at SSRN 3864965*, 2021.
- [51] R. Leonarduzzi, G. Rochette, J.-P. Bouchaud, and S. Mallat, “Maximum-entropy scattering models for financial time series,” in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 5496–5500.
- [52] L. Gatys, A. S. Ecker, and M. Bethge, “Texture synthesis using convolutional neural networks,” *Advances in neural information processing systems*, vol. 28, 2015.
- [53] O. Lanford, “Time evolution of large classical systems,” *In Dynamical systems, theory and applications*, pp. 1–111, 1975.