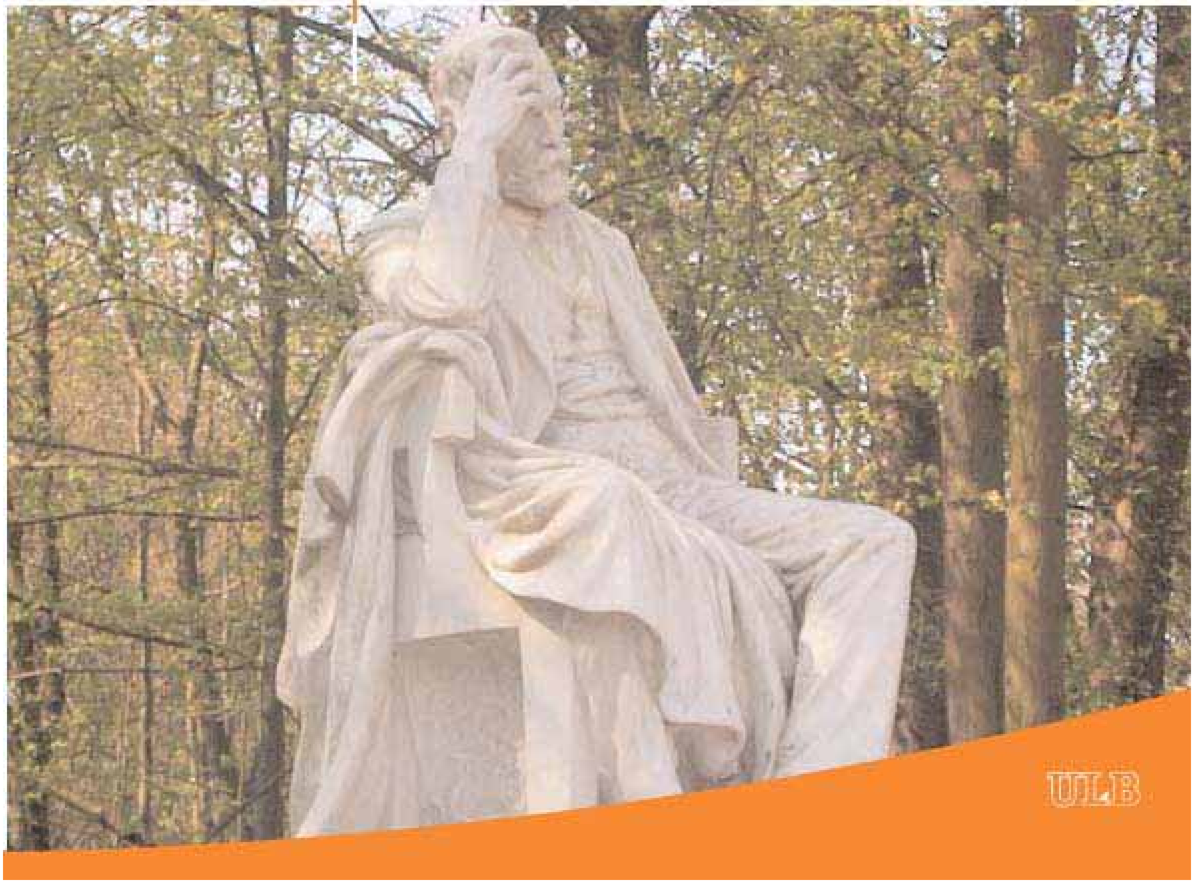
Université Libre de Bruxelles – Solvay Business School – Centre Emile Bernheim
ULB CP145/1 50, Av. F.D. Roosevelt 1050 Bruxelles - BELGIUM

Solvay Business School

Centre Emile Bernheim
Research Institute in Management Sciences

WORKING PAPER: WP-CEB 05-009



Estimation of the Stylized Facts of a Stochastic Cascade Model

Celine Azizieh*
Université Libre de Bruxelles
Belgium

Wolfgang Breymann†,
Department of Mathematics
ETHZ
CH – 8092 Zürich
Switzerland

Abstract

We present a time series model that integrates properties from Lévy-type and multifractal models. Formally it is a stochastic volatility model with discrete time steps, t-distributed return innovations and a stochastic cascade for the volatility process. This model reproduces very well different stylized facts which cannot be reproduced together by other classes of models. We also present an estimation procedure based on the reproduction of stylized facts. This procedure is general and can easily be adapted and/or extended to other models. It may be considered as an extension of the generalized method of moments.

*Chargée de recherche au Fonds National de la Recherche Scientifique

†Research supported by Credit Suisse Group, Swiss Re, and UBS AG through RiskLab, Switzerland.
Present address: University of Basel, Basel, Switzerland

1 Introduction

Statistical properties of return data in finance have been extensively studied during the last years. Currently, properties like leptokurtosis, stochastic volatility effects, occurrence of extremes, seasonalities, and scaling behaviour are well established for different time horizons in the univariate case and refered to as stylized facts. Another well known empirical phenomenon is the so called smile effect, also linked to leptokurtosis. Many extensions of the Black-Scholes model have been studied in view of reproducing those empirical phenomena. Still, it is difficult to reproduce all these properties at all time horizons simultaneously with a single dynamics model. In particular most of the currently favored models are not able to reproduce both the heavy-tailed return distribution at short time horizons and the long-term autocorrelation structure of absolute returns. This, however, is important for applications as hedging and option pricing.

Time series models, especially (G)ARCH models and variants, were extensively used in the eighties and nineties to reproduce the stylized facts, mainly volatility clustering and leptokurtosis. This type of model has the advantage of being more easily estimated, and of working pretty well for daily data and a single time horizon. However, shortcomings show up when intra-day data are used or when several time horizons are considered simultaneously. They have been addressed by various ad-hoc extensions of ARCH-type models, with moderate success.

The strength of (G)ARCH-type models is their relative simplicity, due to the fact that volatility is a *deterministic* function of the previous returns as well as its own past. This is however a severe limitation of modelling, and models involving an additional noise term in the volatility have been considered. The main shortcoming of the corresponding models, so-called (discrete time) stochastic volatility models, was due to the fact that the stochastic term in the volatility makes practically impossible a direct maximum likelihood estimation of the parameters.

In the community of financial mathematics, continuous time models were also used during the last years to model some stylized facts. One of these are the jump-diffusion models, for which discrete versions can also be considered. They are able to reproduce almost every return distribution at a fixed time horizon but they fail to reproduce the interrelation between time horizons, which mainly manifests in scaling behavior.

Another approach are multifractal processes. The problem of this type of models is the fact that the process of interest is in general not a semi-martingale. By construction they reproduce well scaling behavior and volatility autocorrelation but they fail in reproducing well the tails of the return distributions at all time horizons.

In this paper we propose a model that unifies, in the point of view of the ability to reproduce stylized facts, discrete versions of jump-diffusion and multifractal models. The result is a volatility model that we will call *stochastic cascade model* (SCM). It contains a finite cascade in the volatility process, which is actually a finite-level approximation of a 'pure' continuous multifractal model, as well as heavy-tailed innovations ε_t. From a practical perspective our approach can be justified by the fact that observations are always discrete. We admit, however, that the thorough formulation of the time-continuous limit of the model may not be an easy task.

The model was first presented in [2], but up to now there was no way of estimating the model parameters. The version we present here has only three adjustable parameters and can therefore be considered as very parsimonious. Because of the presence of heavy tailed return innovations and a multiplicative cascade for the volatility process, this model integrates important properties from Lévy-type and multifractal models.

To estimate the parameters we present a method based directly on the stylized facts ('stylized facts estimation method', SFEM). It could be considered in some sense as an extension of the generalized method of moments. We can show that indeed both, heavy tails in returns and the volatility cascade are required to reproduce well the shape of the return distributions at different time horizons, scaling properties and the long range of the volatility autocorrelation.

The paper is outlined as follows. In section 2 we briefly describe the data and revisit the important stylized facts of financial time series. In section 3 we give a brief overview over jump-diffusion models and multifractal models and their discrete approximations. In section 4 the SCM is described into details. An estimation procedure based on the stylized facts is presented in 5 together with some results, and section 6 concludes.

2 Stylized facts revisited

2.1 The data

We investigate a high-frequency price series of USD/DEM spot rates. Before revising the stylized facts the following preliminary steps have been performed: collection and filtering, regularization and transformation to logarithmic middle prices, and deseasonalization. They are described in turn.

2.1.1 Collection and filtering

The data set consists of tick–by–tick data originating mainly from Reuters, collected and filtered by Olsen Data. It consists of a large part of the quotes emitted, but not all since the market coverage of the data providers it not complete and depends on the region of the world. The high-frequency series are irregularly spaced; they start in January 1987 and end in December 1998. A single quote at time t consists of a bid price, p_t^{Bid}, and an ask price, p_t^{Ask}. In a first step the data are cleaned by means of a special filter, described in [7] and that tries to take peculiarities of the financial market into account. Among others it corrects for decimal errors caused by the transmission line and removes automatically generated fake quotes during inactive periods used by market participants to test the transmission channel. Since the filter only removes a small fraction of quotes, the filtered time series is still irregularly spaced, and the total amount of data points is quite high (about 10 million).

2.1.2 Regularization and transformation to logarithmic middle prices

To reduce the data we use linear interpolation to transform the time series into a regularly spaced one with step size δ equal to 5 minutes. Since we are not interested in effects related to the bid–ask spread, we work with middle logarithmic prices x_t defined as

$$x_t = \frac{\log\left(p_t^{Bid} \cdot p_t^{Ask}\right)}{2}. \tag{1}$$

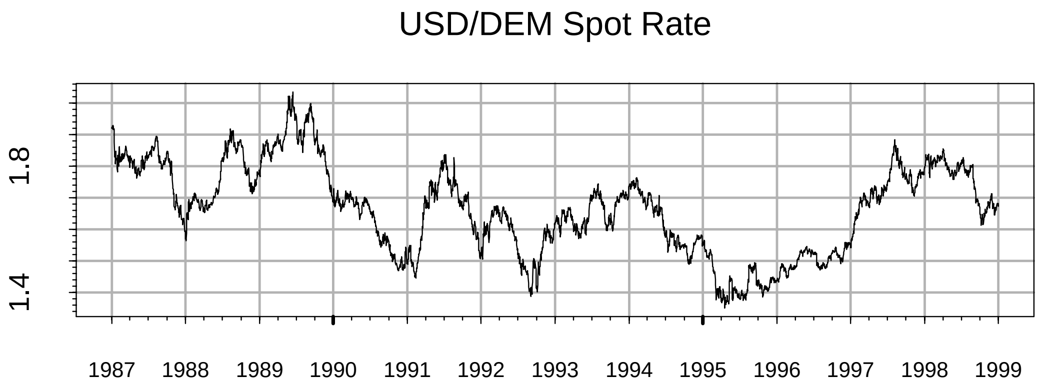


Figure 1: USD/DEM prices from January 1987 to December 1998.

Returns with respect to a time horizon ΔT are then defined as the difference of middle logarithmic prices:

$$r_t[\Delta T] = x_t - x_{t-\Delta T}. \tag{2}$$

2.1.3 Deseasonalisation

Practically any financial time series exhibit seasonalities. The most striking one is the absence of any activity during weekends, which causes a weekly seasonality in the autocorrelation function of lagged absolute returns. With high frequency data the problem of seasonality becomes much more important and more difficult to handle because the entire form of the weekly activity pattern has to be taken into account. In the autocorrelation function of hourly absolute returns, the weekly and daily periods can be distinguished, as shown in [7, 1]. Deseasonalization is done by time transformation. The autocorrelation of absolute returns of the resulting time series decays very smoothly, see Figure 3.

2.2 Heavy tails of the returns distribution

This fact was first observed by Mandelbrot in [14] and Fama in [11] for certain financial time series, and since that time, many models have been proposed to reproduce heavy tailed returns. It is well known that at sufficiently high frequencies, the return distribution becomes heavy tailed, with also a high peakedness at the center as well as a skew to the left. However, due to the symmetric definition of FX rates, the latter is pretty small in the present case. This fact is illustrated in the table below. The kurtosis[1] appears as a decreasing function of the time interval. See also figure 2.

Summary statistics for USD/DEM:

[1] defined here by $\frac{\mathbb{E}[(r(\Delta t)-\mathbb{E}[r(\delta t)])^4]}{\sigma(r(\Delta t))^4]} - 3$, so that the normal distribution has a kurtosis of 0.

Δt	Mean	Median	Variance	Skewness	Kurtosis
1 hour	-2.611331e-006	1.455549e-007	1.249252e-006	1.093956e-002	8.929781e+000
6 hours	-1.554108e-005	-6.716262e-006	7.928494e-006	5.835062e-002	5.580481e+000
24 hours	-6.302301e-005	-2.532633e-005	3.241831e-005	9.915957e-003	2.549666e+000
1 week	-3.220495e-004	-2.901334e-004	1.623375e-004	6.981845e-003	1.051265e+000

The tail index[2] of the returns distribution has been estimated in several papers (see for example [17, 6]), and the conclusion of them was a tail index somewhere between 3 and 4 if $\Delta t \approx 1$ hour.

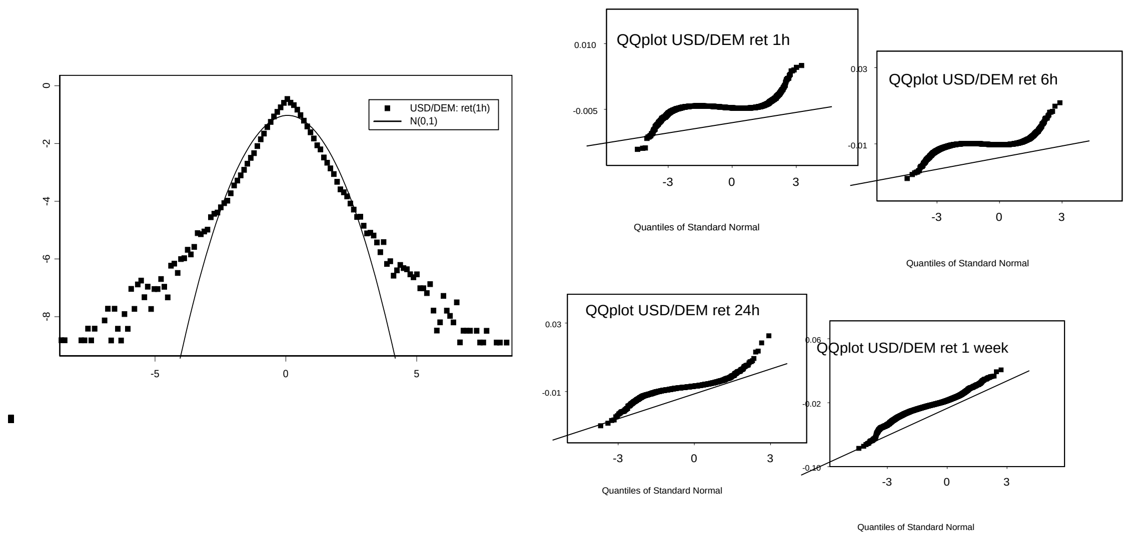


Figure 2: Left: Log density of hourly USD/DEM returns compared with the log density of a standard normally distributed random variable. Right: QQplot of the distribution of returns with respect to a standard normal for different time intervals (1 hour, 6 hours, 1 day and 1 week)

2.3 Volatility clustering – Quasi long range dependence

Although the empirical autocorrelation function (ACF) of returns is consistent with the hypothesis of non autocorrelation, the ACF of absolute or squared returns decays very slowly with respect to lags, as shown on the left graphic in figure 3. The right graphic shows the ACF of absolute returns in a double logarithmic scale, suggesting an autocorrelation function with a hyperbolic tail.

It is common to define the realized volatility at instant t_i for the time interval Δt as:

$$v(t_i) = v(\Delta t, n, p; t_i) := \left(\frac{1}{n} \sum_{j=0}^{n-1} |r(t_{i-j}, \Delta t)|^p \right)^{\frac{1}{p}}$$

where $t_i - t_{i-1} = \Delta t$ and $n \in \mathbb{N}, p \in \mathbb{R}^+, \Delta t \in \mathbb{R}^+$ are fixed. In practice, one chooses $p = 1$ or 2 to have less sensitivity to extreme events.

Commonly, a slowly decreasing ACF is taken as indication for the fact that volatility is clustered, i.e. the process of realized volatilities has not a uniform intensity, but clusters by periods: some of high and some of low volatility.

[2] defined as the order of the highest finite absolute moment

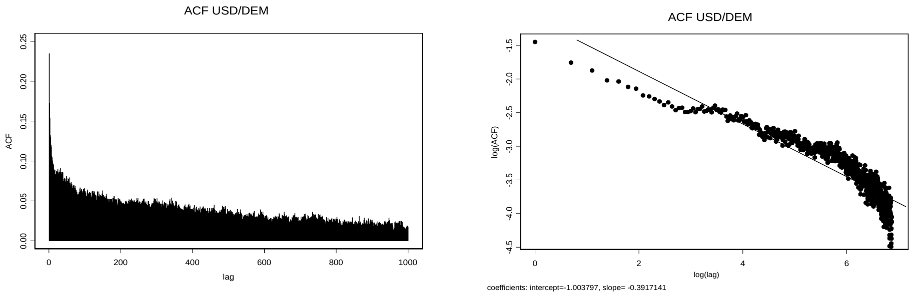


Figure 3: Left: ACF of hourly absolute returns for the data. Right: ACF of hourly absolute returns in a double logarithmic scale for the data. Slope of the regression line: -0.392.

2.4 Scaling properties

First, we define the *structure function* of the returns by

$$S_q(\Delta t) := \frac{\Delta t}{N} \sum_{j=1}^{N/\Delta t} |r(j\Delta t, \Delta t)|^q = \overline{|r(t, \Delta t)|^q}.$$

Then we say that a time series of prices of length N shows scaling properties if

$$S_q(\Delta t) \propto (\Delta t)^{\tau(q)} \tag{3}$$

for some function $\tau(q)$.

If one looks at the behaviour of the structure function of financial data, one typically observes a scaling behaviour with a strictly concave function $\tau(q)$. Such a behaviour is an indication of multifractal property.

We computed $S_q(\Delta t)$ for several values of q and Δt on our data, and the values seem to follow a linear function of $\log(\Delta t)$ showing thus scaling behaviour in very good approximation (see the left graphic on figure 4). We then made a (standard) linear regression for each fixed q and the slopes of the different regression lines were then plotted, giving the values of $\tau(q)$ in (3) (see the right graphic on figure 4).

2.5 Market heterogeneity

The foreign exchange (FX) market is a global over-the-counter market characterized by a geographical repartition of agents having different risk profiles, institutional contraints, etc. The heterogeneous market hypothesis states that agents will react differently to the same information. In [16] the authors focus on the time interval on which investors are acting. The authors compute cross correlations between realized volatilities for different time intervals Δt for different FX rates, and they find some asymmetry in them. They interpret this fact as evidence for the existence of a net information flow from long term traders to short term ones. This could be seen as linked to the so-called feedback effect. Market makers react immediately to new information. If this information is important enough, long term traders will also react, having an effect on the volatility on greater horizons. Finally, a feedback effect appears: market makers react again to the reaction (or absence of reaction) of long term traders, having an impact on the short time volatility.

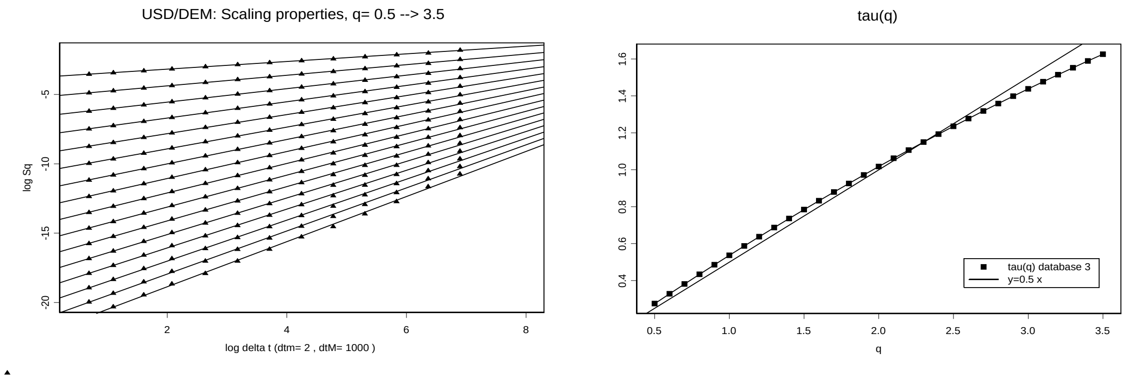


Figure 4: Scaling properties for the data. Left: computation of $S_q(\Delta t)$ for several values of q and Δt on the data. Right: values of the corresponding $\tau(q)$ obtained with a standard linear regression on the values of the left graphic.

3 Modeling approaches

In this section we give a brief overview over two important classes of models, Lévy-type and multifractal models. While Lévy-type models mainly take into account the heavy-tailed return distribution, multifractal models reproduce volatility clustering quite nicely.

3.1 Jump-diffusion processes

A Levy process is a stochastic process that has by definition independent and homogeneous increments. It is well admitted that Brownian motion provides a poor description of the evolution of the (log) prices of financial assets. The situation is improved if one substitutes the Brownian motion by a Lévy process, but to take into account the quasi-long range dependence observed generally in the data, it seemed more convenient to model the volatility itself by such a process. This kind of model can take into account heavy tailedness of the returns as well as quasi long-range dependence. For an overview of the question, consult the works of Barndorff-Nielsen, Shephard, Prause, Eberlein, Keller, Mikosh, Resnick, and the references cited therein.

A discrete approximation of Lévy processes is simply given by

$$r_t = \sigma_t \, \varepsilon_t \tag{4}$$

with heavy-tailed iid-distributed innovations ε_t. The class of distributions of ε_t can be quite general, e.g., generalized hyperbolic distributions [8, 9, 10]. In the SCM model, we restrict ourselves to the simpler case of t-distributions.

3.2 Multifractal processes

Another type of model proposed in financial modeling is multifractal processes. Let us cite the recent works [**?**, 4, 5, 15, 3, 12], where the processes considered contain stochastic multifractality. Actually, the idea of fractal properties of financial data is much older that those works: it has been the starting point of Mandelbrot's seminal investigations.

Here below we present the basic notions about statistical fractality and multifractality.

3.2.1 Definitions

The most striking idea of fractals is self-similarity. For a stochastic process this means qualitatively that at all time scales the fluctuations should look similar after rescaling their intensity. Following the definition of [18] we say that a stochastic process $\{X(t), t \in T\}$ is self-similar with index H if for all $a > 0, n \in \mathbb{N}, t_1, \ldots, t_n \in T$, we have

$$(X(at_1), X(at_2), \ldots, X(at_n)) \stackrel{\mathrm{d}}{=} (a^H X(t_1), a^H X(t_2), \ldots, a^H X(t_n)). \tag{5}$$

Example of self-similar processes are

- Brownian motion: self-similar of index $H = 1/2$.
- Fractional Brownian motion (FBM) of index $H \in (0, 1)$, which is a gaussian process $\{X(t) \,|\, t \in \mathbb{R}^+\}$ with mean zero and covariance function

  $$R_H(t, s) = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |s - t|^{2H})\mathrm{Var}(X(1)). \tag{6}$$

  FBM of index H is self-similar of index H.

In a qualitative way one could define a fractal process as one for which nearly all the trajectories have similar irregularities on all scales. To quantify the irregularity of a deterministic function or a path of a process, one can define several singularity exponents. One of the most popular is the Hölder exponent. More precisely, the local Hölder exponent $h(t)$ at point t of a process X is defined as

$$h(t) := \liminf_{\varepsilon \to 0} \frac{\log_2 \sup_{|u-t|<\varepsilon} |X(u) - X(t)|}{\log_2(2\varepsilon)}. \tag{7}$$

For example, is can be shown that a fractional Brownian motion of index H has nearly all trajectories characterized by $h(t) \equiv H$. Some other exponents can also be introduced (using wavelets or some others more adapted to increasing processes like the case of the cumulative distribution of a random measure, see [18] for details), and they give us a first mean of analyzing the fractal structure of the process.

A multifractal process can be defined as a process whose path singularities are not uniform. More precisely, if one analyses a multifractal process with some chosen exponent $\alpha(t), t \in \mathit{supp}(X)$, one will find that $\alpha(.)$ takes a whole continuum of values and not a single one anymore like in the case of self-similar processes. Note that since we look at stochastic processes, the singularity exponent $\alpha(t)$ will also be stochastic, since it is computed for a fixed trajectory.

3.2.2 Multifractal spectra

Once a singularity exponent $\alpha(.)$ has been chosen, we can define several multifractal spectra, which will quantify which values of the singularity exponent $\alpha(.)$ appear on a trajectory as well as their frequency of appearance. The Hausdorff spectrum of a process X is defined in the following way. For all $a \in \mathbb{R}$, let $E^{(a)} := \{t \,|\, \alpha(t) = a\}$. The *Hausdorff spectrum* is the function

$$\mathbb{R} \to \mathbb{R} : a \mapsto \dim(E^{(a)}) \tag{8}$$

where $\dim(E^{(a)})$ denotes the Hausdorff dimension of the set $E^{(a)} \in \mathbb{R}$. Typically, the support of a multifractal process will be a disjoint union of very complicated sets $E^{(a)}$ having a non-integer fractal dimension. By coming back to the definition of the Hausdorff dimension, one can see that this spectrum gives some information about the frequency of appearance of a value a for the chosen exponent $\alpha(.)$. One can also define other spectra (with other notions of fractal dimension, see [18]), but the idea beyond them will be the same.

One can see that this spectrum (or more precisely a deterministic version of it) is generally linked (by Legendre transform) with the following scaling exponent, defined for a process $X(t)$ ($t \in [0,1]$ without loss of generality):

$$T(q) := \liminf_{n\to\infty} -\frac{1}{n} \log_2 \mathbb{E}_\Omega[S^{(n)}(q)],$$

where

$$S^n(q) := \sum_{k=0}^{2^n-1} \exp\left(-qn\alpha_k^{(n)} \ln(2)\right) = \sum_{k=0}^{2^n-1} 2^{-nq\alpha_k^{(n)}},$$

and α_k^n is the coarse singularity exponent associated to $\alpha(.)$ on dyadic intervals. More precisely, if α is the Hölder exponent h defined previously and $t \in [0,1]$,

$$h_{k_n(t)}^{(n)} := -\frac{1}{n} \log_2 \left(\sup\{|X(u) - X(t)| : u \in [(k_n(t)-1)2^{-n}, (k_n(t)+2)2^{-n})\}\right)$$

where $k_n(t)$ is the unique $k \in \{0,\dots,2^n-1\}$ such that $t \in I_k^n$, with $I_k^n = [k2^{-n},(k+1)2^{-n})$, $n \in \mathbb{N}$. This exponent $T(q)$ has to be put in relation with the exponent $\tau(q)$ obtained in relation (3). One can see on a large amount of examples that typically the multifractality of a process gives rise to a nonlinear $T(q)$.

3.2.3 Multiplicative cascades

A particular and classical case of multifractal functions or processes is given by multiplicative cascades.

Let us first recall the construction of the deterministic binomial cascade. A series of measures μ_i is defined through an iterative scheme, and the binomial measure is then obtained as the limit of infinitely many iterations of this procedure.

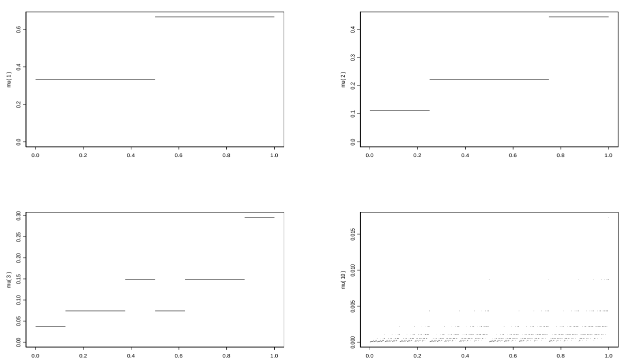


Figure 5: Density of the measures $\mu_1, \mu_2, \mu_3, \mu_{10}$.

One fixes first two real numbers $m_0, m_1 \in (0,1)$ with $m_0 + m_1 = 1$.

At step 1, one defines the measure μ_1 by splitting $[0,1]$ into two intervals $[0,1/2]$ and $[1/2,1]$ and by distributing a mass m_0 uniformly on $[0,1/2]$ and m_1 on $[1/2,1]$. The numbers m_i fulfill the normalisation condition $m_0 + m_1 = 1$.

At the next step we define the measure μ_2 by splitting each of the intervals $[0,1/2]$ and $[1/2,1]$ into two subintervals, $[0,1/4],[1/4,1/2]$ and $[1/2,3/4],[3/4,1]$, and by distributing uniformly on $[0,1/4]$ a fraction m_0 of the mass $\mu_1[0,1/2]$, on $[1/4,1/2]$ a fraction m_1 of $\mu_1[1/1/2]$, and similarly by putting uniformly on $[1/2,3/4]$ a fraction m_0 of $\mu_1[1/2,1]$ and on $[3/4,1]$ a fraction m_1 of $\mu_1[1/2,1]$, that is:

$$\begin{aligned} &\mu_2[0,1/4] = m_0 m_0, &&\mu_2[1/4,1/2] = m_0 m_1, \\ &\mu_2[1/2,3/4] = m_0 m_1, &&\mu_2[3/4,1] = m_1 m_1. \end{aligned}$$

Then we continue this scheme, defining a sequence μ_k of peacewise uniform measures, uniform in fact on each dyadic interval of length 2^{-k}. Finally, if t is of the form

$$t = \sum_{i=1}^{k} \zeta_i 2^{-i}, \qquad \zeta_i \in \{0,1\} \text{ for all } i,$$

for some fixed $k \in \mathbb{N}$, then

$$\mu_{k+j}([t, t+2^{-k}]) = \mu_k([t, t+2^{-k}]) \quad \forall j \in \mathbb{N},$$

and

$$\mu([t, t+2^{-k}]) = m_0^{k-n_1} m_1^{n_1}$$

where $n_1 = \sum_{i=1}^{k} \zeta_i$. Figure 5 illustrates the density of different measures μ_k ($k = 1,2,3,10$). The measure μ is defined as the limit of this sequence (μ_k). By the preceeding, μ is defined for each dyadic interval and consequently, it is well defined on each Borel set. The cumulative distribution function of μ, defined by $\mathcal{M}(t) := \mu([0,t])$, is called a binomial cascade. One can see that μ has no well-defined density anymore, or equivalently, $\mathcal{M}$ is nowhere differentiable.

One can generalize this construction with some chosen $b \in \mathbb{N}$ and positive real numbers $m_0, \ldots, m_b$ with $\sum_{i=0}^{b-1} m_\beta = 1$ and subdivide each interval at a step into b subintervals. One can also randomize the construction by replacing the m_β by random variables M_β with independence between all steps of the construction and $\sum M_\beta = 1$ (conservative cascades) or only in expectation. The corresponding multiplicative cascade $\mathcal{M}(t)$ becomes then a stochastic process. One can see that $\mathcal{M}(t)$ has a multifractal structure, in the sense that its multifractal spectrum will take a hole range of values and not a single one. The SCM described in the following section will contain some kind of generalization of such a random multiplicative measure.

4 The stochastic cascade model

The stochastic cascade model introduced in [2] is motivated by an analogy between turbulence and finance proposed in [13], and tries to take into account the different stylized facts described above. It will, on the contrary of both preceeding models, be expressed in discrete time, in the tradition of econometricians.

The study of turbulence phenomena consists in the analysis of the velocity field of fluids with a high Reynolds number in certain domains. Fluids having a sufficiently high Reynolds number will show important and unstable fluctuations. At very high Reynolds numbers, the fluid will show eddies at all spacial scales. One important phenomenon is the so-called Richardson cascade: mechanical energy is introduced into the fluid at large spacial scales, large eddies begin to appear, will soon break up into smaller eddies, splitting themselves into smaller ones and so on, until a limit spacial scale at which energy is dissipating into heat.

If one makes the following parallelism:

$$\begin{array}{rcl} \text{energy} & \leftrightarrow & \text{realized volatility} \\ \text{distance} & \leftrightarrow & \text{time delay} \\ \text{velocity at point } x & \leftrightarrow & \text{price at instant } t \end{array}$$

then the statistical properties of the series of interest show some similarities, and the Richardson cascade can be put in parallel with the net information flow from long to short time horizons.

Let us recall now the stochastic cascade model from [2]. One considers a cascade of time horizons $\tau^{(1)} > \cdots > \tau^{(m)}$, representing the different time horizons of traders, where $\tau^{(1)}$ is typically of the order of one year and $\tau^{(m)}$ of the order of a few minutes. We denote the logarithmic return at time t by $r_t := r_t = r(t, \Delta t = \tau^{(m)})$ where $t = t_0 + n\tau^{(m)}$ for some fixed initial instant t_0 and $n \in \mathbb{N}$.

We then suppose that

$$r_t = \sigma_t \xi_t$$

where ξ_t are i.i.d. random variables $\sim t_N$, Student-t with N degrees of freedom. The main difference between an ARCH-type model and this one is that the volatility σ_t is not autoregressive anymore but is a multiplicative cascade as described below. We assign to each time horizon a level of volatility $\sigma_t^{(k)}$ which will be determined by the volatility at level $k-1$ as well as by a stochastic factor $a_t^{(k)}$:

$$\sigma_t^{(k)} = a_t^{(k)} \sigma_t^{(k-1)}.$$

We then arrive to a volatility of the form:

$$\sigma_t^{(m)} = \sigma_t = \sigma_0 \prod_{k=1}^{m} a_t^{(k)},$$

where we suppose that σ_0 is a constant and $\{a_t^{(k)}, k = 1, \ldots, m\}$ are random variables following some renewal process described below and taking into account the net information flow from long to short time horizons.

We also suppose for simplicity that

$$\frac{\tau^{(k)}}{\tau^{(k-1)}} = p$$

for some constant $p < 1$.

Let us now describe the form of the different factors appearing in the volatility. We suppose first that at the initial instant t_0, all the variables $a_t^{(k)}$ are generated following independent log-normal probability distributions $LN(a^{(k)}, \lambda_k^2)$ (where $a^{(k)} = \mathbb{E}[\log(a_{t_0}^{(k)})]$ and $\lambda_k^2 = \mathrm{Var}\left[\log(a_{t_0}^{(k)})\right]$.) Then, to pass from instant t_n to $t_{n+1} = t_n + \tau^{(m)}$,

$$a_{t_{n+1}}^{(1)} = \begin{cases} a_{t_n}^{(1)} & \text{with probability } 1 - \omega^{(1)} \\ \text{renewed as an independent } LN(a^{(1)}, \lambda_1^2) & \text{with probability } \omega^{(1)} \end{cases}$$

for some renewal probability $\omega^{(1)}$ that will be chosen later. In case of renewal of $a_{t_{n+1}}^{(1)}$, we suppose that $a_{t_{n+1}}^{(k)}$ is also renewed for all $k > 1$ following independent $LN(a^{(k)}, \lambda_k^2)$. In the other case (no renewal), one supposes that

$$a_{t_{n+1}}^{(2)} = \begin{cases} a_{t_n}^{(2)} & \text{with probability } 1 - \omega^{(2)} \\ \text{renewed as an independent } LN(a^{(2)}, \lambda_2^2) & \text{with probability } \omega^{(2)} \end{cases}$$

and so on. The renewal probabilities $\omega^{(k)}$ are chosen such that the expected time interval between two renewals of $a_t^{(k)}$ is equal to $\tau^{(k)}$, as described below.

By construction, the number of renewals of $a_t^{(1)}$ is a Bernouilli process (with time units equal to $\tau^{(m)}$) of probability $\omega^{(1)}$. It is well known that for such a process, the success instants (here the renewal instants) T_k follow a negative binomial law and the intervals between two consecutive successes are i.i.d. geometric of parameter $\omega^{(1)}$. This implies that the expected time interval between two renewals, $\mathbb{E}[T_k - T_{k-1}]$, is equal to $1/\omega^{(1)}$. If we want this expectation be equal to $\tau^{(1)}/\tau^{(m)}$ time units $\tau^{(m)}$, since moreover $\tau^{(1)}/\tau^{(m)} = p^{1-m}$, one has to impose

$$\omega^{(1)} = p^{m-1}.$$

Let us now consider the renewals of $a_t^{(2)}$. If we denote by $S_n^{(2)}$ the number of renewals after n steps (after n time intervals of length $\tau^{(m)}$), and if we denote by $\xi_n^{(1)}$ the indicating random variable of a renewal at step n of the process $a_t^{(1)}$, one has

$$\begin{aligned} \mathbb{P}[S_n^{(2)} - S_{n-1}^{(2)} = 1] &= \mathbb{E}[\mathbb{P}[S_n^{(2)} - S_{n-1}^{(2)} = 1 \,|\, \xi_n^{(1)}]] \\ &= \omega^{(1)} + \omega^{(2)}(1 - \omega^{(1)}) \\ &\stackrel{\text{def}}{=} \tilde{\omega}^{(2)}. \end{aligned}$$

Since the renewals of $a_t^{(1)}$ are independent of time, it will be the same for those of the $a_t^{(2)}$. The variables $S_n^{(2)} - S_{n-1}^{(2)}$ are thus i.i.d. Bernouilli random variables with parameter $\tilde{\omega}^{(2)}$, and the number of renewals with respect to time is again a Bernouilli process of parameters $\tilde{\omega}^{(2)}$. The expected time interval between two renewals is hence given by $1/\tilde{\omega}^{(2)}$, and if we want it to be equal to $\tau^{(2)}/\tau^{(m)} = p^{2-m}$, we have then to impose

$$\omega^{(1)} + \omega^{(2)}(1 - \omega^{(1)}) = p^{m-2},$$

that is

$$\omega^{(2)} = \frac{p^{m-2} - p^{m-1}}{1 - p^{m-1}}.$$

One can repeat this reasoning for the following steps of the cascade. More generally, the renewal probability of the process $a_t^{(k)}$ at a given instant will be given by

$$\tilde{\omega}^{(k)} = \tilde{\omega}^{(k-1)} + \omega^{(k)}(1 - \tilde{\omega}^{(k-1)}),$$

and we will choose $\tilde{\omega}^{(k)} = p^{m-k} = \tau^{(m)}/\tau^{(k)}$, which implies

$$\omega^{(k)} = \frac{p^{m-k} - p^{m-k+1}}{1 - p^{m-k+1}}, \quad k = 2, \ldots, m.$$

The construction of the stochastic volatility σ_t is similar to that of a multiplicative measure described above, except that the different factors of the cascade keep their value during the corresponding time interval only in mean. So this is like a multiplicative measure in mean. Another important difference is the fact that one stops the cascade at an elementary time scale $\tau^{(m)}$. This has the advantage that in a discrete time setting one can work with (non-gaussian) heavy-tailed innovations. It will be shown in the next section that this is really necessary.

5 Estimation of the model parameters

We tried to make an estimation of some parameters in view of the three aspects:

- scaling properties;
- autocorrelation of absolute returns;
- distribution of returns.

Here below is explained the stylized facts estimation method. For each stylized fact, one considered some distance function. The idea is the following: we calculated the adequate quantity for some amount of simulated series, computed the mean value over those different series, taking also into account the volatility of the quantity calculated on simulated data. We then took the difference with the corresponding quantity calculated on real data.

5.1 The parameters

Since we worked with hourly data, we choose to fix $\tau^{(m)} = 0.5h$ and $\tau^{(1)} \sim 1$ year as well as $p = 1/\sqrt{2}$. We only selected then one point on two in the simulated series, meaning that a simulation of length $n_2 = 200000$ gives only 100000 selected points used for further analysis, and corresponding to hourly ticks. It seemed that the choice of p was not so crucial. In what follows, we fixed the value of the parameter p since it seemed that it had few influence on the results. On the contrary, we made vary the number of degree of freedom N in the Student innovations ξ_i, as well as some parameters c_1, c_2 chosen in the construction of $\Lambda^2 = \sum_{i=1}^{m} \lambda_i^2$, i.e. some parameters linked with the volatility of the stochastic volatility. More precisely, the parameters λ_k have been generated in the following way. If we note $\gamma := -\log(p)$, we suppose that

$$\lambda_k^2 = \gamma c_1 + \gamma^2 k c_2$$

for some parameters c_1, c_2. This assumption is motivated by empirical findings [13]. It implies the following form for the variance of the logarithm of σ_t:

$$\Lambda^2 = \sum_{i=0}^{m} \lambda_i^2 = \gamma m c_1 + \frac{\gamma^2}{2}(m^2 - m)c_2.$$

We then made vary the parameters c_1 and c_2. We also supposed that σ_0 is a constant.

5.2 Method of estimation

For each stylized fact mentioned above, we constructed a distance function to be minimized. We simulated n_1 time series of length n_2 (we choose in practice $n_1 = 20$, $n_2 = 200000$). For each fact and each simulated time series j, we computed a matrix of values $(x_{i,j})_{i\in I, j=1\ldots n_1}$ (where I is some index set depending on the stylized fact considered) taken as estimates of certain quantities linked with the stylized fact itself. Beside this vector of values, a matrix of weights $(\omega_{i,j})_{i\in I}$ (with $\sum_{j=1}^{n_1} \omega_{i,j} = 1$) was computed at the same time, reflecting the importance (or credibility) to be given to every component of the vector $(x_{i,j})_{i\in I}$. A vector of mean values over the n_1 simulations is then obtained as $\bar{x}_i := \sum_{j=1}^{n_1} x_{i,j}\omega_{i,j}$, and a final vector of weights is obtained as the inverse of the square roots of the following weighted variances: $\bar{w}_i := \sum_{j=1}^{n_1} x_{i,j}^2 \omega_{i,j} - (\bar{x}_j)^2$.

Scaling properties: Concerning scaling properties, since no analytical computation could be made (the cascade being stopped at the level $\tau^{(m)}$ and the innovations being non gaussian), we computed empirically the scaling exponents for $q = 0.5$ to 3.5 by steps of 0.5. The method is the following. For each simulated time series, we calculated for each fixed q the structure function $S_q(\Delta t)$ (as defined in Section 2.4) for different values of Δt (equidistant in a logarithmic scale, from 1 to 512) and made a weighted linear fit of $\log(S_q(\Delta t))$ over $\log(\Delta t)$ with weights equal to $S_q(\Delta t)/\sqrt{Var[|r(\Delta t)|^q]}$ (here $Var[|r(\Delta t)|^q]$ denotes the sample variance of the absolute returns with time interval Δt to the power q along a simulated time series). We then obtain for each simulation j and each q an estimation $\tau_j(q)$ of $\tau(q)$ by taking the slope of the corresponding regression line. This is the choice of the matrix $(x_{i,j})$. We then take, at q fixed again, the weighted mean of those estimations $\tau_j(q)$ over the n_1 simulations, weighted by the respective inverse of the corresponding standard error in the different linear fits (this is the choice of the matrix of weights $w_{i,j}$).

Volatility clustering: Concerning this fact, $x_{i,j}$ was simply chosen as the ACF of absolute returns, with a maximum lag of 1000 (so here the vector $x_{\cdot}, j$ has a length of 1000). The weights $w_{i,j}$ were chosen constant (i.e. $w_{i,j} = 1/n_1$ for all i, j).

Distribution of log returns: We then computed the empirical cumulative distribution functions of returns on different time intervals, so not only for one interval. This is interesting because the behaviour of those different functions is also linked with multifractality and scaling behaviour. For each computation of the distribution, we calculated it on a grid of 2000 points. We gave more weight to the points in the tails than to those in the center of the distribution. This was achieved by introducing, for a fixed simulation, some weight function proportional to $\omega_k = n^2/((n-k+1)*k)$, where "k" counts the points

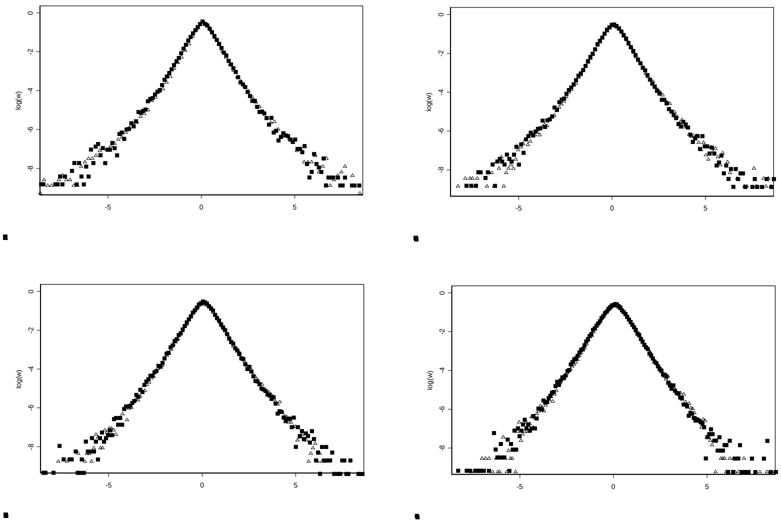


Figure 6: Log density of the returns for different time intervals ($\Delta t = 1, 3, 6, 10$, from the left to the right) for data (black squares) and simulations of the model (white triangles) with N=4 (the number of degrees of freedom), $c_1 = 0.03$ and $c_2 = -0.001$.

in the ordered sample of length n (in ascending order). This means that the median of the distribution corresponds to $k = n/2$. The weight function is of course normalized such that $\sum \omega_k = 1$. The length n depends on the time horizon Δt considered, so that n implicitly depends on Δt. Concerning the weights associated to the different simulations ('$w_{i,j}$'), we chose them constant.

The distance function of each stylized fact is then obtained as the Euclidian norm of the weighted difference between the mean values vector $(\bar{x}_i)_{i \in I}$ over the simulations and the corresponding values over the observations $(y_i)_{i \in I}$ (calculated in the same way as $(x_{i,j})$ but with $n_1 = 1$ and $j \in \{1\}$), that is

$$f := (\sum_{i \in I} |\bar{x}_i - y_i|^2 \bar{\omega}_i)^{1/2}.$$

The total distance is then obtained as the (possibly weighted) sum of those distance functions.

5.3 Results

We found a quite similar behaviour for the distances associated with long range dependence and the distribution of returns. The sensitivity to different simulations of both functions was quite reasonable, in the sense that the variation of those two functions with respect to the three parameters c_1, c_2, N was relatively smooth. On the contrary, the distance associated with the scaling properties was quite unstable, it is the function for which we found the largest fluctuations. That is precisely on this stylized fact that there is disagreement in the scientific community. This means that it seems better to replace this function by the distance associated with the distribution of returns for different time intervals. This last function also captures multifractality, and is apparently more stable.

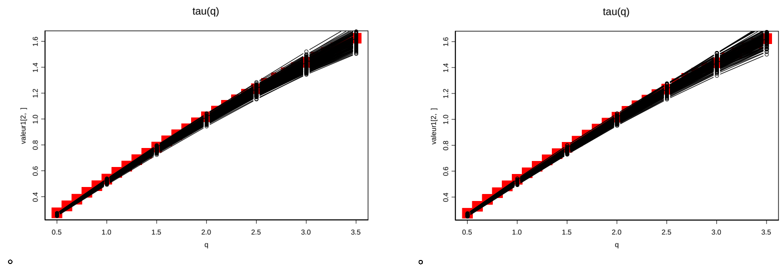


Figure 7: Scaling exponents for the data and 100 simulated series of the model with N=4, $c_1 = 0.03$ and $c_2 = -0.001$ (left) and $c_1 = 0.025$ and $c_2 = -0.0005$ (right).

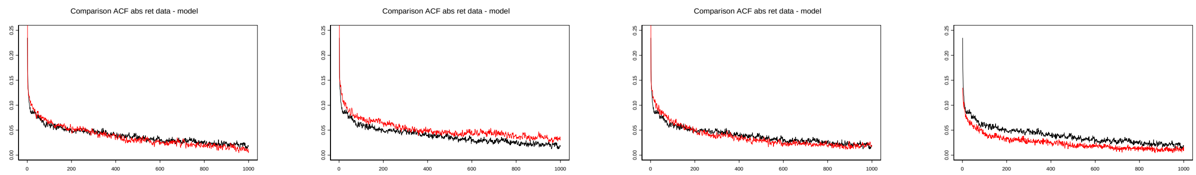


Figure 8: ACF of absolute returns for the data (in black) and different simulations of the model (red) (N=4, $c_1 = 0.025$, $c_2 = -0.0005$).

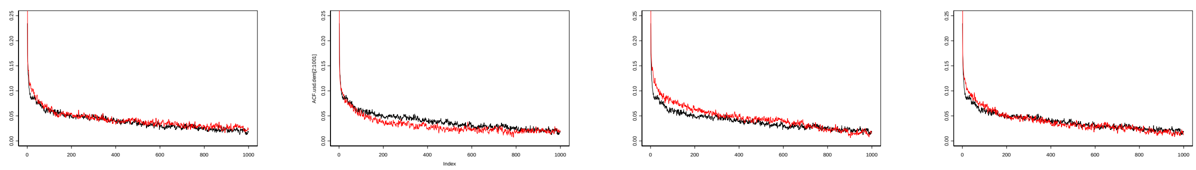


Figure 9: ACF of absolute returns for the data (in black) and different simulations of the model (in red) (N=4, $c_1 = 0.03$, $c_2 = -0.001$).

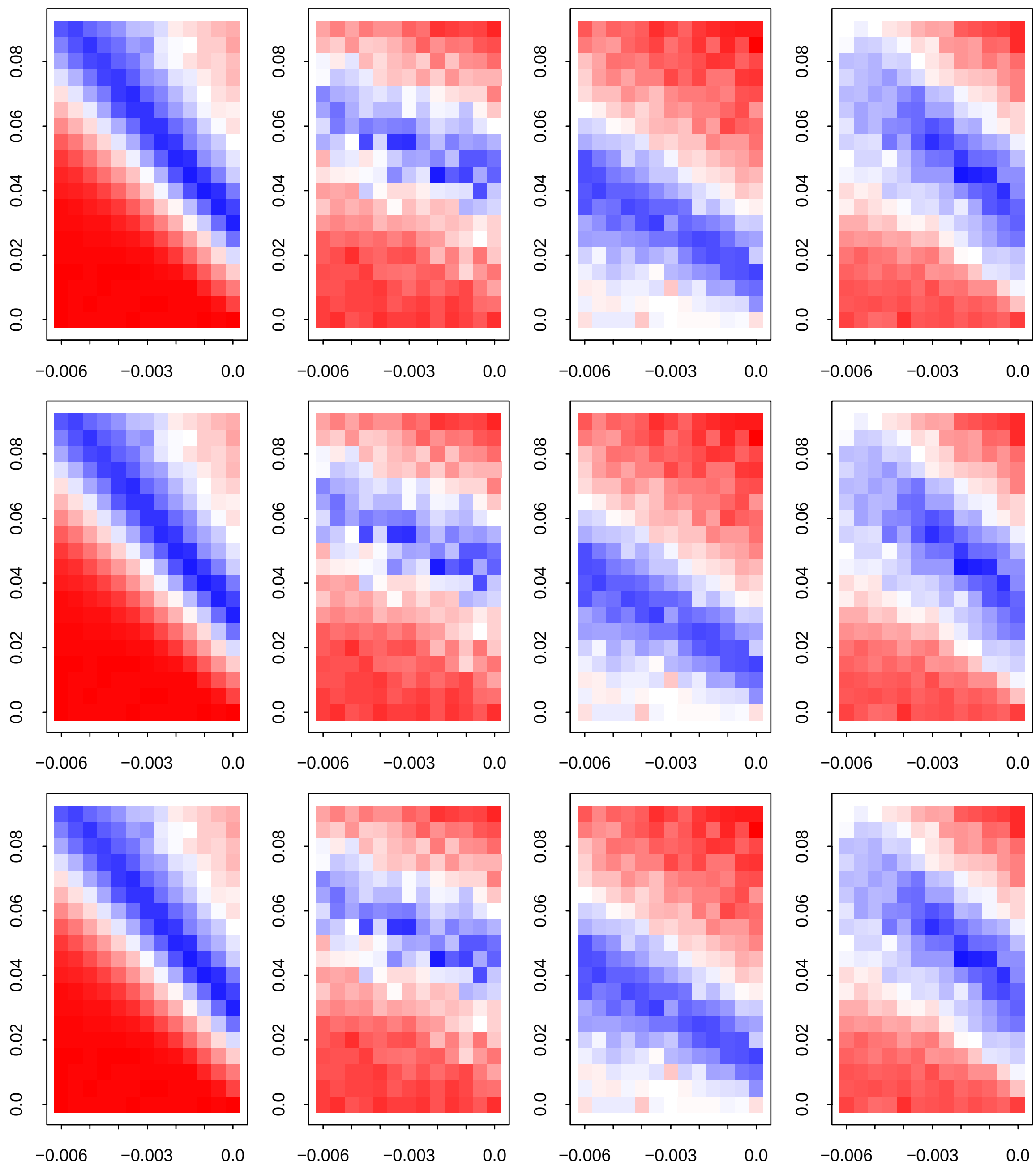


Figure 10: Color-coded density plots of the distance function. The columns represent the distance functions for ACF (left), scaling (middle left), distributions (middle right) and total (right). The number of degrees of freedom of the t-distribution varies from 3.4 (top), to 3.8 (bottom) in steps of 0.2. In each panel the x-axis represents c_2 and the y-axis c_1.

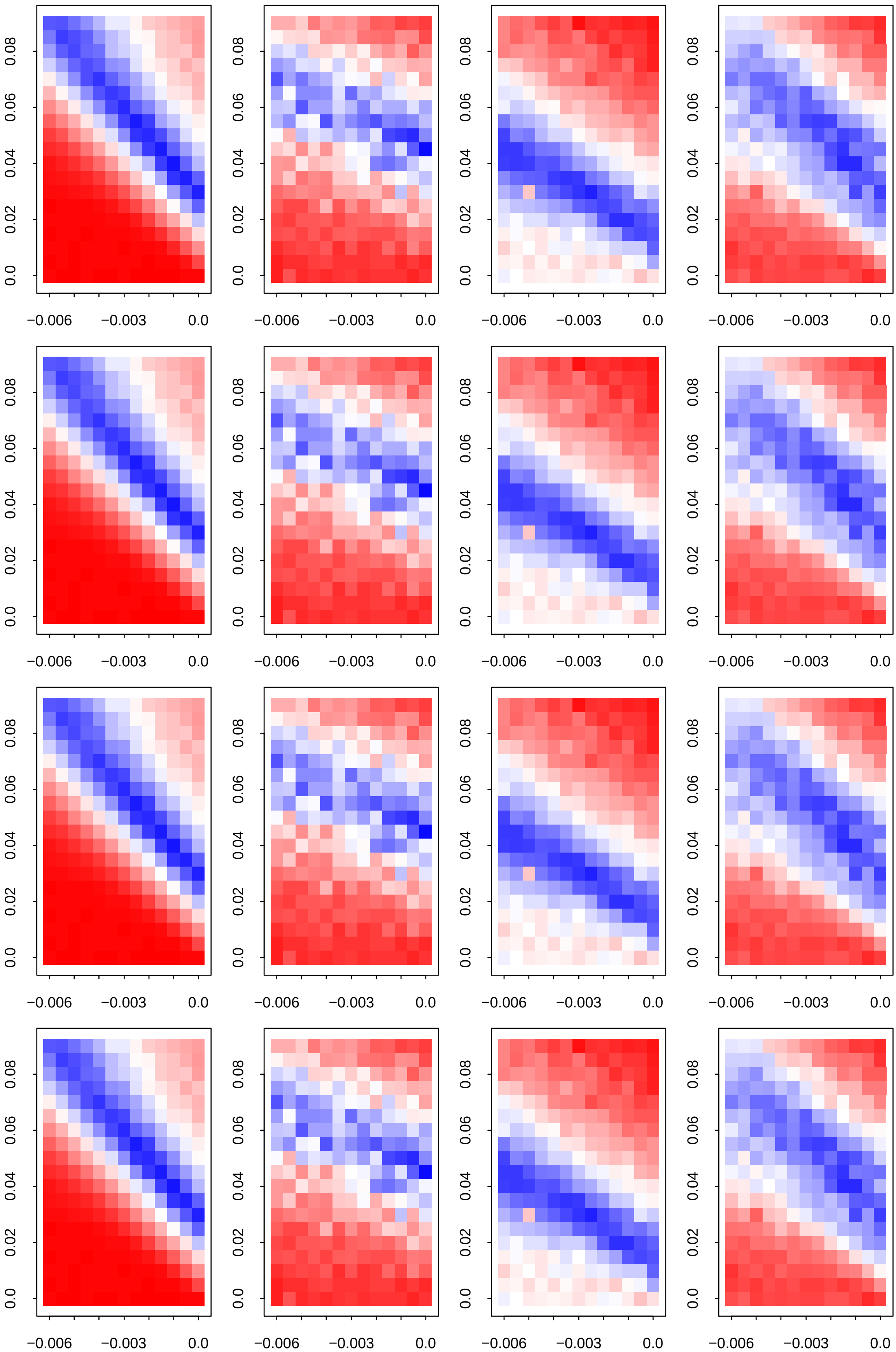


Figure 11: Same as fig. 10 with number N of degrees of freedom varying from 4 (top) to 4.6 (bottom).

This also means that the present dispute over scaling properties between members of the scientific community is non well-funded because it relies on an unstable quantity.

Figures 6–9 illustrate the adequacy of the model in vue of the different stylized facts considered.

Figures 10 and 11 presented in this paper give an illustration of the graphs of the three types of distance functions we considered to adjust the parameters c_1, c_2 and N. We made them vary as described in the table below:

	Minimum value	Maximum value	Step
c_1	0	0.09	0.005
c_2	-0.006	-0.002	0.0005
N	3.4	4.6	0.2

Those values were chosen in that way after having experimented on many simulations the sensibility of each distance function and located the zone of the parameters giving the best estimates.

The figures show contour plots for each distance function, in a separate column, left column for the ACF function of absolute returns, middle left column for scaling properties, middle right for distributions as well as right column for some aggregation of the three functions. The blue zones correspond to smaller values of the distance functions, and red zones to higher values.

Each row corresponds to a fixed value of the number of degrees of freedom N (see caption), the two other parameters c_1, c_2 varying in their respective set of values.

As explained before, we see in general that the graphs of 'ACF' and 'distribution' depend quite smoothly on the parameters, while the distance function relative to scaling laws shows some instability, even with a large number of simulations.

This suggests that such a method of estimation, taking into account the distribution as well as the decay of the ACF of absolute returns is quite promising, while an estimation based on scaling properties seems very difficult. This is interesting since usually multifractality has been captured with the study of scaling properties. However, the distribution of returns on different time intervals also captures the multifractal behaviour, and seems a more stable way of estimation. Thus this is an alternative to take into account multifractality in the fit of the parameters.
It seems also that the parameter N has less influence on the estimation of the optimal value than c_1, c_2, since the optimal region evolves quite similarly in each row. In fact, for a chosen 'reasonable' value of N, it is always possible to adapt sufficiently well c_1, c_2 to have a quite good fit.

6 Conclusion

After revisiting the stylized facts, we recall and precise in more details the Stochastic Cascade Model proposed in [2]. This model tries to take into account three statistical properties observed usually on a lot of financial data, that is heavy tailedness of the distribution of log returns, with a tail index depending on the interval considered for the returns, quasi long range dependence (quantified by the ACF function of absolute returns) as well as nonlinear scaling laws. We also propose a method of estimation of some parameters of the model based on the stylized facts themselves, called stylized facts estimation methods. The results obtained suggest that such a method of estimation, taking into account the distribution of returns for different time intervals as well as the decay of the ACF of absolute returns is quite promising, while an estimation based on scaling properties seems very difficult. So the distribution of returns on different intervals appears as a better alternative to the use of scaling law for the estimation of the parameters.

The next step will be the construction of a parametric test for the values of the parameters c_1, c_2.

References

[1] Breymann, W., A. Dias, & P. Embrechts (2003). Dependence structures for multivariate high-frequency data in finance. *Quant. Finance.* **3**(1), 1–14.

[2] W. Breymann, S. Ghashghaie et P. Talkner, A stochastic cascade model for FX dynamics, *Journal of Theoretical and Applied Finance*, **3**, (2000) 357-360.

[3] L. Calvet, A. Fisher, B.B. Mandelbrot, Large deviations and the distribution of price changes, Cowles Fondation Discussion Paper # 1165 (1997).

[4] L. Calvet, A. Fisher, Forecasting Multifractal Volatility, Journal of Econometrics 105, 27-58 (2001).

[5] L. Calvet, A. Fisher, Catastrophic Risk in Multifractal Asset Returns,? with L. Calvet in Risk and Return Management for Insurance Companies, ed. I. Vanderhoof and E. Altman, Kluwer Academic Press, forthcoming.

[6] Rama Cont, Statistical Finance: Empirical and Theoretical Approaches to the Statistical Modelling of Price Variation in Speculative Markets, Ph. D. thesis, CMAP-Ecole Polytechnique (1998).

[7] M. Dacorogna, R. Gençay, U. A. Müller, R. B. Olsen, O. V. Pictet, An introduction to high frequency finance, Academic Press, 2001.

[8] E. Eberlein, Application of generalized hyperbolic Levy motion to finance, in Barndorff-Nielsen, O.E., Mikosh, T & Resnick S. (Eds): Levy processes - theory and applications, Boston: Birkhauser (2000).

[9] E. Eberlein and U. Keller, Hyperbolic distributions in finance, *Bernouilli* **1**, 281-299 (1995).

[10] E. Eberlein and S. Raible, Term structure models driven by general Levy processes, *Math. Finance* **9**, 31-54 (1999).

[11] E.F. Fama, The behaviour of stock markets prices, *Journal of Business*, **38**(1), 34-105 (1965).

[12] A. Fisher, L. Calvet, B.B. Mandelbrot, Multifractality of Deutschemark/US Dollar exchange rates, Cowles Fondation Discussion Paper Ï165 (1997).

[13] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, Y. Dodge, Turbulent cascades in foreign exchange markets, *Nature* **381**, 776-770 (1996).

[14] B.B. Mandelbrot, The variation of certain speculative prices, *Journal of Business*, **36**, (1963), 394-419.

[15] B.B. Mandelbrot, A. Fisher, L. Calvet, The Multifractal Model of Asset Returns, Cowles Fondation for Economic Research Working Paper (1997).

[16] U. A. Müller, M.M. Dacorogna, R.D. Davé, R.B. Olsen, O.V. Pictet,J.E. von Weizsächer, Volatilities of different time resolutions – Analysing the dynamics of market components, *Journal of Empirical Finance* **4** (1997) 213-239.

[17] O.V. Pictet, M.M. Dacorogna, U.A. Müller, Hill, bootstrap and jackknife estimators for heavy tails, dans *A practical guide to heavy tails: statistical techniques for analysing heavy tailed distributions,* R.J. Adler, R.E. Feldman and M.S. Taqqu, Eds., pp. 283-310.

[18] R.H. Riedi, Multifractal processes, preprint to be published.