Minding Impacting Events in a Model of Stochastic Variance

Sílvio M. Duarte Queirós[1]*[¤], Evaldo M. F. Curado[2], Fernando D. Nobre[2]

1 Centro de Física do Porto, Porto, Portugal, 2 Centro Brasileiro de Pesquisas Físicas and National Institute of Science and Technology for Complex Systems, Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

We introduce a generalization of the well-known ARCH process, widely used for generating uncorrelated stochastic time series with long-term non-Gaussian distributions and long-lasting correlations in the (instantaneous) standard deviation exhibiting a clustering profile. Specifically, inspired by the fact that in a variety of systems impacting events are hardly forgot, we split the process into two different regimes: a first one for regular periods where the average volatility of the fluctuations within a certain period of time W is below a certain threshold, ϕ, and another one when the local standard deviation outnumbers ϕ. In the former situation we use standard rules for heteroscedastic processes whereas in the latter case the system starts recalling past values that surpassed the threshold. Our results show that for appropriate parameter values the model is able to provide fat tailed probability density functions and strong persistence of the instantaneous variance characterized by large values of the Hurst exponent ($H > 0.8$), which are ubiquitous features in complex systems.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sdqueiro@gmail.com

¤ Current address: Istituto dei Sistemi Complessi - CNR, Roma, Italy

Introduction

For the last years the physical community has broaden its subject goals to matters that some decades ago were too distant from the classical topics of Physics. Despite being apparently at odds with the standard motivations of Physics, this new trend has given an invaluable contribution toward a more connected way of making Science, thus leading to a better understanding of the world surrounding us [1]. Within this context, the major contribution of physicists is perhaps the quantitative procedure, reminiscent of experimental physics, in which a model is proposed after a series of studies that pave the way to a reliable theory. This path has resulted in a series of findings which have helped such diverse fields as physiology, sociology and economics, among many others [2–4]. Along these findings, one can mention the determination of non-Gaussian distributions and long-lasting (power-law like) correlations [5–7]. Actually, by changing the observable, the conjunction of the two previous empirical verifications is quite omnipresent. For this reason and regardless the realm of the problem very similar models have been applied with particular notoriety to discrete stochastic processes of time-dependent variance based on autoregressive conditional heteroscedastic models [8]. That is to say, most of these models are devised taking basically into account the general features one aims at reproducing, rather than putting in elements that represent the idiosyncracies of the system one is surveying. For instance, many of the proposals cast aside the cognitive essence prevailing on many of these systems, when it is well known that in real situations this represents a key element of the process [9]. On the other hand, intending to describe long-lasting correlations, long-lasting memories are usually introduced thus neglecting the fact that we do not traditionally keep in mind every happening. As a simple example, we are skilled at remembering quotidian events for some period. However, we will discard that information as time goes by, unless the specific deed either created an impact on us or has to do with something that has really touched us somehow. In this case, it is likely that the fact will be remembered forever and called back in similar or related conditions, which many times lead to a collective memory effect [10].

In this work, we make use of the celebrated heteroscedastic model, the $ARCH$ process [11] and modify it by pitching at accommodating cognitive traits that lead to different behavior for periods of high agitation or impact. Particularly, we want to stress on the fact that people tend to recall important periods, no matter when they took place. To that end, we introduce a measure of the local volatility, as well as a volatility threshold, so that the system changes from a normal dynamics, in which it uses the previous values of the variable to determine its next value, to a situation in which it recalls the past and compares the current state with previous states of high volatility, even if this past is far.

Standard models of heteroscedasticity

The Engle's formulation of an autoregressive conditional heteroscedastic ($ARCH$) time series [11] represents one of the simplest and effectual models in Economics and Finance, for which he was laureated the Nobel Memorial Prize in Economical

Sciences in 2003 [12]. Explicitly, the *ARCH* corresponds to a discrete time, t, process associated with a variable, z_t,

$$z_t = \sigma_t \, \omega_t, \qquad (1)$$

with ω_t being an independent and identically distributed random variable with zero mean and standard deviation equal to one. The quantity σ_t represents the time-dependent standard deviation, which we will henceforth name *instantaneous volatility* for mere historical reasons. Traditionally, a Gaussian is assigned to the random variable ω_t, but other distributions, namely the truncated α-stable Lévy distribution and the q-Gaussian (Student-t) have been successfully introduced as well [13,14]. In his seminal paper, Engle suggested that the values of σ_t^2 could be obtained from a linear function of past squared values of z_t,

$$\sigma_t^2 = a + \sum_{i=1}^{s} b_i \, z_{t-i}^2, \qquad (a, b_i \geq 0). \qquad (2)$$

In financial practice, *viz.*, price fluctuations modelling, the case $s = 1$ ($b_1 \equiv b$) represents the very most studied and applied of all the *ARCH(s)*-like processes. The model has been often applied in cases where it is assumed that the variance of the observable (or its fluctuation) is a function of the magnitudes of the previous occurrences. In a financial perspective, Engle's proposal has been associated with the relation between the market activity and the deviations from the normal level of volatility a, and the previous price fluctuations making use of the impact function [8]. Alternatively, recent studies convey the thesis that leverage can be responsible for the volatility clustering and fat tails in finance [15]. Nonetheless, the heteroscedastic *ARCH*-like processes has been repeatedly used as a forecasting method. In other words, one makes use of the magnitude of previous events in order to indicate (or at least to bound) the upcoming event (see e.g. [16,17]). In respect of its statistical features, although the time series is completely uncorrelated, $\langle z_t \, z_{t'} \rangle \sim \delta_{tt'}$, it can be easily verified that the covariance $\langle |z_t| \, |z_{t'}| \rangle$ is not proportional to $\delta_{tt'}$. As a matter of fact, for $s = 1$, it is provable that $\langle z_t^2 \, z_{t'}^2 \rangle$ decays according to an exponential law with a characteristic time $\tau \equiv |\ln b|^{-1}$. This dependence does not reproduce most of the empirical evidences, particularly those bearing on price fluctuations studies. In addition, the introduction of a large value of s used to give rise to implementation problems [18]. Expressly, large values of s augment the difficulty of finding the appropriate set of parameters $\{b_i\}$ for the problem under study as it corresponds to the evaluation of a large number of fitting parameters. Aiming to solve this short-coming of the original *ARCH(1)* process, the *GARCH(s,r)* process was introduced [19] (where G stands for generalized), with Eq. (2) being replaced by,

$$\sigma_t^2 = a + \sum_{i=1}^{s} b_i z_{t-i}^2 + \sum_{i=1}^{r} c_i \, \sigma_{t-i}^2 \qquad (a, b_i, c_i \geq 0). \qquad (3)$$

In spite of the fact that the condition, $b + c < 1$, guarantees that the *GARCH(1,1)* process exactly corresponds to an infinite-order *ARCH* process, an exponential decay for $\langle z_t^2 z_{t'}^2 \rangle$, with $\tau \equiv |\ln(b + c)|^{-1}$ is found.

Although the instantaneous volatility is time dependent, the *ARCH(1)* process is actually stationary with the *stationary variance* given by,

$$\langle \sigma^2 \rangle = \widehat{\sigma^2} = \frac{a}{1-b}, \qquad (b < 1), \qquad (4)$$

(herein $\langle \ldots \rangle$ represents averages over samples at a specified time and $\widehat{\ldots}$ denotes averages over time in a single sample). Moreover, it presents a stationary probability density function (PDF), $P(z)$, with a kurtosis larger than the kurtosis of distribution $P(\omega)$. Namely, the fourth-order moment is,

$$\langle z^4 \rangle = a^2 \langle \omega^4 \rangle \frac{1+b}{(1-b)(1-b^2 \langle \sigma^4 \rangle)}.$$

This kurtosis excess is precisely the outcome of the dependence of σ_t on the time (through z). Correspondingly, when $b = 0$, the process is reduced to generating a signal with the same PDF of ω, but with a standard variation equal to $\sqrt{a}$. At this point, it is convenient to say that, for the time being and despite several efforts, there are only analytical expressions describing the tail behavior of $P(z)$ or the continuous-time approximation of the *ARCH(1)* process with the full analytical formula still unknown [14,20].

In order to cope with the long-lasting correlations and other features such as the asymmetry of the distribution and the leverage effect, different versions of the *ARCH* process have been proposed [8,18]. To the best of our knowledge, every of them solve the issue of the long-lasting correlations of the volatility by way of introducing an eternal dependence on z_t^2 in Eq. (2), $b_i \equiv b \mathcal{K}(i)$, with $\mathcal{K}(.)$ representing a slowly decaying function [8,21]. Most of these generalizations can be encompassed within the fractionally integrated class of *ARCH* processes, the *FIARCH* [22–25]. The idea supporting the introduction of a power-law for the functional form of $\mathcal{K}(.)$ is generally based on the assumption that the agents in the market make use of exponential functions $\mathcal{K}(.)$ with a broad distribution of relaxation times related to different investment horizons [26,27]. This type of model has achieved a huge popularity in the replication of non-Gaussian time series in several areas, such as biomedicine, climate, engineering, and physics (a few examples can be found in [28–35]).

As described above, the statistical features of the macroscopic observables are the result of the nature of the interactions between the microscopic elements of the system and the relation between microscopic as well as the macroscopic observables. In the case of the "financial" *ARCH* process, it was held that z_i^2 bears upon the impact of the price fluctuations on the trading activity. On the one hand, it is understood that the impact of the price fluctuations (or trading activity) on the volatility does not merely come from recent price fluctuations and it does actually involve past price fluctuations. In finance, upgraded versions of heteroscedasticity models use multi-scaling, *i.e.*, it is assumed that the price will evolve by modulating the volatility according to the volatility over different scales (days, weeks, months, years, etc.) [36] in order to smooth their possible misjudgement about the volatility. However, in practice, these models do not differ much from *FIARCH*-like proposals at the level of the results we are pointing at. Alternately, it is worthwhile to look upon the *ARCH* proposal as a mechanism of forecast [16,17]. In this way, the simplest approach, the *ARCH(1)*, represents an attempt to foresee future values just taking into account recent observations, whereas models like the *FIARCH* bear in mind all the history weighting each past-value according to some kernel functional.

Minding impacting events

In our case, we want to emphasize the fact that people tend to recall periods of high volatility (*i.e.*, impact) in the system, no

matter when they took place, by changing the surrounding conditions as agent-based models suggested [37,38]. Hence, we introduce a measure of the local volatility,

$$v_t = \frac{1}{W} \sum_{i=0}^{W-1} z_{t-i}^2, \tag{5}$$

and a threshold, ϕ, so that instead of Eq. (2), the updating of σ_t^2 goes as follows:

$$\sigma_t^2 = \begin{cases} a + \sum_{i=1}^{t} b_i\, z_{t-i}^2 & \text{if } v_{t-1} < \phi, \\ a + \sum_{i=1}^{t} b_i'\, z_{t-i}^2 & \text{if } v_{t-1} \geq \phi, \end{cases} \tag{6}$$

where $b_i = b\mathcal{K}(i) = b \exp[-i/\tau]$ [39,40]. Therefore, if we assume the financial market perspective, we are implicitly presuming that the characteristic time, τ, is Dirac delta or at least narrow distributed, so that the exponential functional is a valid approximation. This approach is confirmed by recent heuristic studies in which it has been verified that the largest stake of the market capitalization is managed by a small number of companies that apply very similar strategies [41]. With the second branch equation we intend to highlight the difference in behavior of the "normal" periods of trading and the periods of significant volatility, in which the future depends on the spells of significant volatility in the past as well. The values b_i' are defined as,

$$b_i' = b p_i \Theta[v_{t-i} - \phi], \tag{7}$$

with $\Theta[\ldots]$ being the Heaviside function and p_i is a factor that represents a measure of the similarity (in the volatility space) between the windows of size W with upper limits at z_t and z_{t-W+1}, respectively. Analytically, this is equivalent to mapping segments in the form $\{z_i^2, \ldots, z_{i-W+1}^2\}$ into vectors in $\mathbb{R}_0^{+W}$ and afterward computing a normalized internal product-like weight,

$$p_i = \frac{1}{\mathcal{N}} \sum_{j=1}^{W} z_{t-j}^2 z_{i-j}^2, \tag{8}$$

where, for the sake of simplicity, we set aside the time dependence of p_i and b_i' in the equations, while $\mathcal{N}$ represents the normalization factor such that $\sum_i p_i = 1$ for all i (with fixed t).

We are therefore dealing with a model characterized by 5 parameters, namely: a (the normal level of volatility) and b (the impact of the observable in the volatility), which were both first introduced by Engle in [11]; τ, put forward in exponential models; and two new parameters W (representing the volatility spell) and ϕ that we will reduce to a single extra parameter. If we think of trading activities, our proposal introduces a key parameter, the volatility threshold, ϕ, which signals a change in behavior of the agents in the market. At present, significant stake of the trading in financial markets is dominated by short-term positions and thus a good part of the dynamics of price fluctuations can be described by Eq. (2), or by functions with an exponential kernel. As soon as the market fluctuates excessively, *i.e.*, the volatility soars beyond the threshold, the market changes its trading dynamics. The main forecast references are obviously the periods where the volatility has reached high levels and afterward, the periods of those which are most similar; this is the rationale described by our Eq. (8).

Thence, our proposal is nothing but the use of simple mechanisms that in a coarse-grained way master a good part of our decisions.

Results

General results

In this section we present the results obtained by the numerical implementation of the model. For comparison, we will use the results of a prior model that can be enclosed in the class of *FIARCH* processes [25]. There, the adjustment of the parameters comes from the delicate balance between the parameter b, which is responsible for introducing deviations of the volatility from its normal level a, and the parameter controlling the memory. On the one hand, large memory has the inconvenient effect of turning constant the instantaneous volatility, so that after a seemly number of time steps the value of σ becomes constant, hence leading to a Gaussian (or close to it) distribution of the variable z, independently of how large b is. On the other hand, short memory is unable to introduce long-range correlations in the volatility, although it enhances larger values of kurtosis excess. The model we introduce herein is rather more complex. In order to deal with the change of regime, we define a parameter establishing this alteration and we need to specify W and τ. Henceforth, we have assumed $W = \tau$, which is very reasonable as it imposes that the volatility and the time scale that the agents in the market use to assess the evolution of the observable are the same. In order to speed up our numerical implementation, we have imposed a cut-off of $10W$ in the computation of the first line in Eq. (6). This approximation turns the numerical procedure much lighter with a negligible effect because the influence of the discarded past is not much relevant in numerical terms (within standard numerical implementation error). In all of our realizations, we have used a normalized level of expected volatility, $a = 1$, and we have defined the volatility threshold in units of $a/(1-b)$, following a stationary approach, as well.

We have adjusted the probability distributions of z by means of the distribution,

$$P(z) = \mathcal{Z}^{-1} \left(1 + B z^{2v}\right)^{\frac{1}{1-q'}}, \tag{9}$$

the behavior of which follows a power-law distribution for large $|z|$ with an exponent equal to $\frac{2v}{q'-1}$ and where (using Ref. [42], sec. 3.194),

$$\int z^n (1 + B z^{2v})^{\frac{1}{1-q'}} dz =$$

$$\frac{1+(-1)^n}{1+n} \frac{\Gamma\left[\frac{2v+n+1}{2v}\right] \Gamma\left[\frac{2v+(n+1)\left(1-q'\right)}{2v(q'-1)}\right]}{B^{\frac{1+n}{2v}} \Gamma\left[\frac{1}{q'-1}\right]}, \tag{10}$$

$\left(\frac{2v}{q'-1} > 1+n\right)$, and $\mathcal{Z}$ represents the previous integral with $n = 0$. The fittings for the probability density distribution (9) were obtained using non-linear and maximum log-likelihood numerical procedures and the tail exponents double-checked with the value given by the Hill estimator [43,44]. As a matter of fact, values of v different from 1 have only been perceived for large values of b and small values of ϕ (slightly larger) or large values of ϕ (slightly

smaller). For $v=1$ and $q' \neq 1$, the PDF corresponds to a q'-Gaussian distribution (or Student-t distribution) [45] and when $q'=1$ we have either the Gaussian ($v=1$) or the stretched distribution ($v \neq 1$). Since that in the majority of the applications one is interested in the tail behavior, we have opted for following the same approach by defining the tail index as,

$$\frac{2}{q-1} = \frac{2v}{q'-1} \Leftrightarrow q = v^{-1}\left(q'+v-1\right). \qquad (11)$$

In spite of the fact that other functional forms could have been used, we have decided on Eq. (9) because of its statistical relevance and simplicity (in comparison with other candidates involving special functions, namely the hypergeometric). Moreover, the q-Gaussian (t-Student) is intimately associated with the long-term distribution of heteroscedastic variables since it results in the exact distribution when the volatility follows an inverse-Gamma distribution [35,46–48].

Concerning the persistence of the volatility, we have settled on the Detrended Fluctuation Analysis (DFA) [49], which describes the scaling of a fluctuation function related to the average aggregated variance over segments of a time series of size ℓ,

$$F(\ell) \sim \ell^H, \qquad (12)$$

where H is the Hurst exponent. Although it has been shown that Fluctuation Analysis methods can introduce meaningful errors in the Lévy regime [50], we have verified that for our case, which stands within the finite second-order moment domain, the results of DFA are so reliable as other scaling methods.

Let us now present our results for $b=0.5$, which is able to depict the qualitative behavior of the model for small b. This case corresponds to a situation of little deviation from the Gaussian, when long-range memory is considered. In accordance, we can analyse the influence of the threshold ϕ and W. Overall, we verify a very sparse deviation from the Gaussian. Keeping W fixed and varying ϕ, we understand that for small values of ϕ the distribution of z_t is Gaussian and the Hurst exponent of $|z_t|$ is 1/2. It is not hard to grasp this observation if we take into account that, by using small values of ϕ, we are basically employing almost all of the past values which limits the values of instantaneous volatility to a constant value after a transient time. As we increase the value of ϕ, we let the dynamics be more flexible and therefore the volatility is able to fluctuate, resulting in a kurtosis excess. For small values of W, the Hurst exponent is slenderly different from 1/2 and the value of the Hurst exponent increases with W. However, because of the small value of b, the rise of W turns out the distribution of z barely undistinguishable from a Gaussian. This behavior is described in Fig. 1. We have obtained a Gaussian distribution and a Hurst exponent $H=0.5$ for small values of ϕ ($\phi=0.1$) and W ($W=5$). When we augment the value of the threshold, $\phi=5$, the system is loose and the instantaneous volatility is able to fluctuate leading to the emergence of tails ($q=1.09$) and a subtle increase of the Hurst exponent ($H=0.52\pm0.01$). Hiking up both W and ϕ ($W=75$ and $\phi=2$), we have achieved large values of the Hurst exponent ($H=0.58\pm0.02$), but the small value of b is not sufficient to induce relevant fluctuations, bringing on a distribution that is almost Gaussian ($q=1.02$). The distribution fittings were assessed by computing the critical value $P_{KS}^* = 1-\alpha_{crit}$ from the Kolmogorov-Smirnov test [51] that are equal to 0.9634 and 0.9454, respectively.

As we increase the value of b, we favor the contribution of the past values of the price dynamics, thus, for the same value of W we

are capable of achieving larger values of the kurtosis excess, that we represent by means of the increase of the q index. The same occurs for the Hurst exponent. This general scenery is illustrated in Fig. 2 for the value $b=0.998$, where we present the dependence of q and H with ϕ, for different choices of W. Again, the higher W, the lower the tail index q, because the extension of the memory surges a weakening of the fluctuations in the volatility. The opposite occurs with the Hurst exponent, which increases towards unit (ballistic regime) as we consider W larger, for obvious reasons. In all the cases of (b,W) investigated, we verified that both q and H augment with ϕ. The assessment of the numerical adjustments is provided in Tab. 1 in the form of the P_{KS}^* critical values from the Kolmogorov-Smirnov test [51]. The only case we obtained a value 1 (within a five-digit precision) was for the pair $W=10$ and $\phi=5$, which results in a value quite close to the limit of finite second-order moment (a fat-tailed distribution with $q=5/3$). At this point it is worth saying that we have investigated the likelihood of other well-known continuous distributions, such as the stretched-exponential, the simple t-Student, Lévy, and Gaussian. Nonetheless, the fittings carried with Eq. (9) outperformed every other analyzed distribution.

Concerning the instantaneous volatility, σ_t, we verified that the Dirac delta distribution, $p(\sigma)=\delta(\sigma-1)$, starts misshaping and short tails appear as we depict in Fig. 3 (upper panel) for the case $b=0.998$, $W=75$ and $\phi=0.25$. Considering this particular case, we can present relevant evidence of the effectiveness of our proposed probability distribution approach. The empirical distribution function in the upper panel of Fig. 3 may be simply approximated by

$$p(\sigma) = \begin{cases} f\dfrac{1}{2c} & \text{if} \quad \sigma \neq 1 \\ (1-f)\delta(\sigma-1) & \text{otherwise} \end{cases}, \qquad (13)$$

with $c \geq 0, f \leq 1$, and $\sigma \in (1-c,1+c)$; when $f=0$ we recover the homoscedastic process distribution as a particular case. Reminding that at each time step the distribution is a Gaussian (conditioned to a time-dependent value of σ) the long-term distribution is,

$$P(z) = \int_{1-c}^{1+c} p(\sigma)\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{z^2}{2\sigma^2}\right]d\sigma, \qquad (14)$$

which gives (Ref. [42], sec. 3.351),

$$\begin{aligned} P(z) = {} & \frac{f}{4\sqrt{2\pi}c}\left(\mathrm{Ei}\left[-\frac{z^2}{2(1-c)^2}\right] - \mathrm{Ei}\left[-\frac{z^2}{2(1+c)^2}\right]\right) \\ & + \frac{1-f}{\sqrt{2\pi}}\exp\left[-\frac{z^2}{2}\right] \end{aligned} \qquad (15)$$

where Ei[.] is the Exponential Integral function (see e.g. Ref. [52]). Considering $c=1/2$ (which is appropriate to the case shown) and taking for the sake of simplicity $f=1/2$, we obtain the function presented in Fig. 4, the kurtosis of which is $\kappa = \dfrac{10854}{3125} \approx 3.47$ (making use of Ref. [42], sec. 5.221). Actually, this curve is represented in the scaled variable z/σ so that the standard deviation, which is originally equal to $\dfrac{(c+1)^3+(c-1)^3+6c}{12c}$, becomes equal to one, like in other depicted distributions. The accordance between this distribution and the empirical distribution is quite remarkable since it emerges from no numerical adjustment and can be further improved by tuning the values
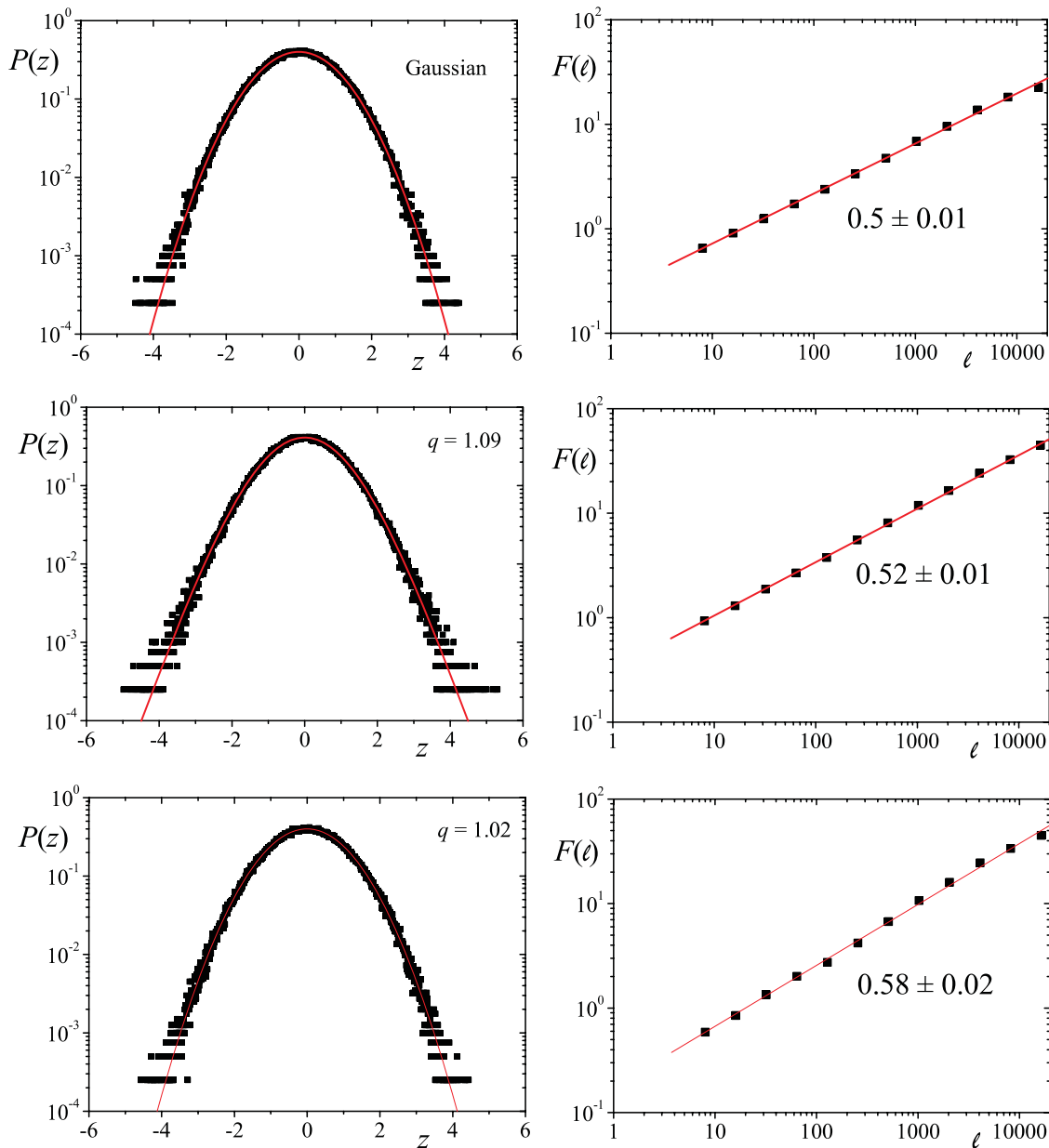
Figure 1. Probability density functions $P(z)$ vs z in a log-linear scale on the left column; On the right column the fluctuation functions $F(\ell)$ vs ℓ for $|z|$ in a log-log scale. The values of the model parameters are: $\phi = 0.1, W = 5$ yielding $q = 1$ and $H = 0.5 \pm 0.01$ (upper panels); $\phi = 5, W = 5$ yielding $q = 1.09 \pm 0.01$ and $H = 0.52 \pm 0.01$ (middle panels); $\phi = 2, W = 75$ yielding $q = 1.02 \pm 0.01$ and $H = 0.58 \pm 0.02$ (lower panels). The results have been obtained from series of 4×10^5 elements and the numerical adjustment of $P(z)$ gave values of χ^2/n never greater than 0.00003, with R never smaller than 0.998.
doi:10.1371/journal.pone.0018149.g001

of f and c. Regardless, this kurtosis value is only 2.2% larger than our numerical adjustment (see Table 1 for the goodness of fitting). Furthermore, comparing the distributions by means of the symmetrized Kullback-Leibler divergence $KL = \frac{1}{2}$ $\left(\int P(z) \ln \frac{P(z)}{P'(z)} dz + \int P'(z) \ln \frac{P'(z)}{P(z)} dz \right)$, we obtain a value of 0.00014 that is 19 times smaller than the distance between our fitting and a Gaussian. These results show that the PDF of Eq. (9) not only provides a good description of the data, but it is much more manageable as well.

Cases for which the kurtosis excess is relevant ($q > 5/4$) stem from wider distributions of σ (see the lower panel of Fig. 3). Actually, it is the emergence of larger values of the instantaneous volatility that brings forth fat tails. Although we have not been successful in describing the whole distribution, we have verified that, for values of $q > 5/4$, the distribution $p(\sigma)$ is very well described by a type-2 Gumbel distribution,

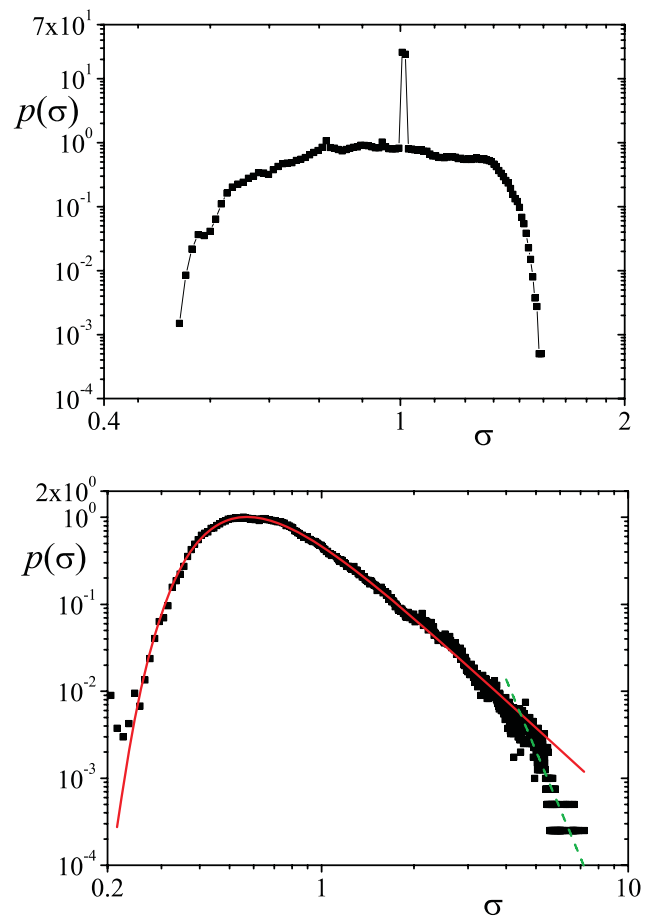$$p(\sigma) \propto \exp\left[-\beta \sigma^{-\zeta}\right] \sigma^{-\zeta - 1}, \qquad (16)$$

and after certain value of σ the distribution sharply decreases
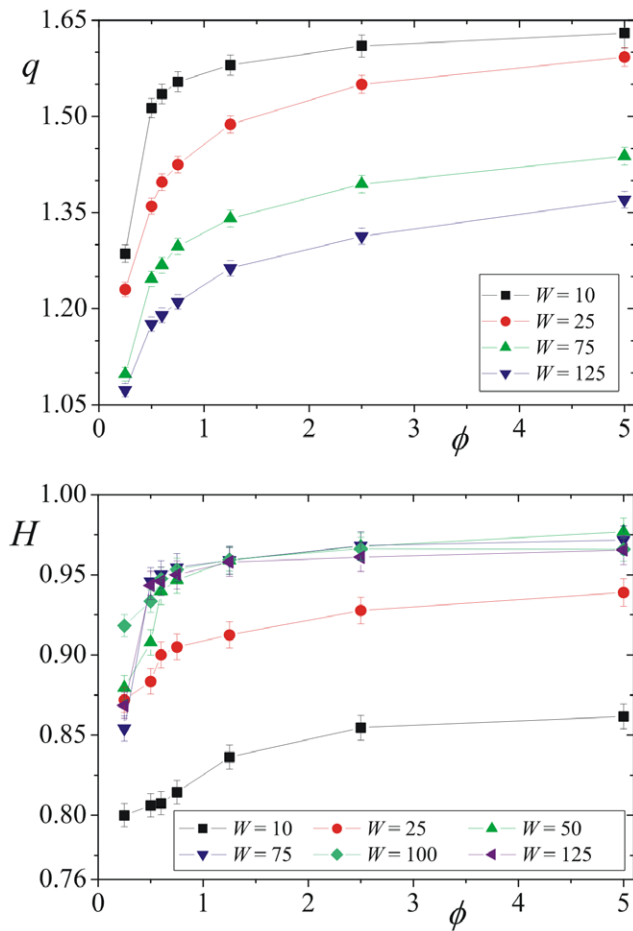
Figure 2. Value of the tail index q vs parameter ϕ for several values of W and $b=0.998$ according to the adjustment procedures mentioned in the text in the upper panel. In the lower panel Hurst exponent H vs ϕ. The results have been obtained from series of 4×10^5 elements and the numerical adjustment of $P(z)$ gave values of χ^2/n never greater than 0.00003 with R^2 never smaller than 0.998. Regarding the values of the Hurst exponent, the absolute error has never been greater that 0.015 and a linear coefficient $R > 0.999$.
doi:10.1371/journal.pone.0018149.g002

according to a power-law with a large exponent. We credit this sheer fall to the threshold ϕ, which introduces a sharp change in the dynamical regime of the volatility and thus in its statistics. In finance, such a cut-off is more than plausible as real markets do suspend trading when large price fluctuations occur. This also grants feasibility to descriptions based on truncated power-law distributions [6]. Moreover, a fall off is also presented in the quantity σ_e of Fig. 3 in Ref. [53]. It is known that for heteroscedastic models the tail behavior of the long-term distribution is governed by the asymptotic limit of $p(\sigma)$ when σ tends to infinity. For the case of distribution (16), this limit is the power-law $\sigma^{-\zeta-1}$ and therefore we can verify that the asymptotic behavior of the long-term distribution of the variable z,

$$
\begin{aligned}
\lim_{|z| \to \infty} P(z) &\sim \int \left[\lim_{\sigma \to \infty} p(\sigma) \right] \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{z^2}{2\sigma^2} \right] d\sigma \\
&\sim \int \sigma^{-\zeta-2} \exp\left[-\frac{z^2}{2\sigma^2} \right] d\sigma,
\end{aligned}
\tag{17}
$$

Figure 3. Probability density function of the instantaneous volatility $p(\sigma)$ vs σ for two different $b=0.998$. In the upper panel: $b=0.998$, $\phi=0.25$ and $W=75$ which leads to a sharply peaked distribution around $\sigma=1$ and to a $P(z)$ tail index $q=1.1$. In the lower panel: $b=0.998$, $\phi=2.5$ and $W=25$ that results in a broader distribution largely described by a type-2 Gumbel distribution with $\beta=0.421\pm0.002$ and $\zeta=2.323\pm0.006$ ($\chi^2/n=0.00011$ and $R^2=0.9982$). For $\sigma \sim 5$, $p(\sigma)$ changes its behavior to a faster decay with an exponent equal to 8.4 ± 0.2 represented by the gray symbols. The ANOVA test of the type-2 Gumbel adjustment (up to $\sigma \sim 5$) have yielded a sum of squares of 0.03553 (323 degrees of freedom) and 20.3684 (2 degrees of freedom) for the error and the model, respectively. The uncorrected sum of the sum of squares is 20.4039 (325 degrees of freedom) and the corrected total is 12.5941 (324 degrees of freedom). The empirical distribution function has been obtained from series of 4×10^5 elements.
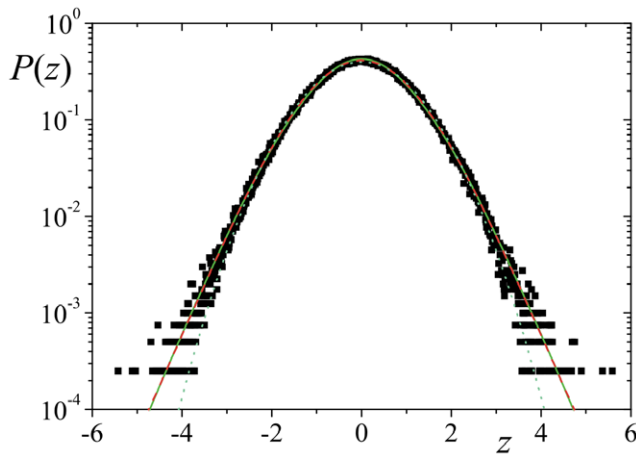doi:10.1371/journal.pone.0018149.g003

yields a power-law distribution (applying Ref. [42], sec. 3.326),

$$
P(z) \to |z|^{-\zeta}, \qquad (z \to \infty). \tag{18}
$$

For $p(\sigma)$ following an exponential decay in the form $\exp[-\gamma\sigma]$, a similar procedure yields,

$$
P(z) \to G\left[\frac{\gamma^2}{8} z^2 \Big|\, {}^{-}_{0,} \ {}^{-}_{\frac{1}{2},} \ {}^{-}_{1} \right], \qquad (z \to \infty), \tag{19}
$$

where $G[.]$ is the Meijer G-function [42,52]. It is worth noting that in an effort to obtain a full description of $p(\sigma)$ we also used a function such as $f(x) = Z \exp\left[-\beta x^{-\zeta}\right] \left(1 - \frac{A}{B} + \frac{A}{B}\exp\left[\frac{A}{\mu}x\right] \right)^{-\mu}$ which

Figure 4. Probability density function $P(z)$ vs z. The points represent the empirical distribution function for $b=0.998$, $\phi=0.25$ and $W=75$; the dashed red line is our adjustment with Eq. (9) with $q=1.1\pm0.01$, $v=1$ and $B=(q-1)/(5-3q)$ [$\chi^2/n=0.00003$ and $R^2=0.9986$]; the green line is PDF (15) with $f=c=1/2$ and the dotted cyan line is the Normal distribution.
doi:10.1371/journal.pone.0018149.g004

allows the appearance of a crossover from a power law to an exponential decay. Nonetheless, it did not provide better results.

It is worth saying that we can reduce the number of parameters to a, b and ϕ, i.e., apply the simple $ARCH(1)$ process, and obtain fat tails and persistence still.

Comparison with a real system

Following this picture, we can now look for a set of parameters that enable us to replicate a historic series such as the daily (adjusted) log-index fluctuations, $\{r(t)\}$, of the SP500 stock index, $\{S(t)\}$, between 3rd January 1950 and 12th April 2010 (14380 data points) with,

Table 1. Critical values $P^*_{KS}=1-\alpha_{crit}$ from the Kolmogorov-Smirnov test for typical pairs (W,ϕ) used for adjustments.

W	ϕ	P^*_{KS}	W	ϕ	P^*_{KS}
10	0.25	0.9997	75	0.25	0.9865
	0.5	0.9998		0.5	0.9898
	0.6	0.9998		0.6	0.9902
	0.75	0.9998		0.75	0.9908
	1.25	0.9999		1.25	0.9918
	2.5	0.9999		2.5	0.9925
	5	1		5	0.9943
25	0.25	0.9985	125	0.25	0.9749
	0.5	0.9989		0.5	0.9761
	0.6	0.999		0.6	0.976
	0.75	0.9991		0.75	0.9767
	1.25	0.9992		1.25	0.9780
	2.5	0.9994		2.5	0.9817
	5	0.9996		5	0.9870

doi:10.1371/journal.pone.0018149.t001

$$r(t)=\ln S(t+1)-\ln S(t). \qquad (20)$$

The adjusted values of the index take into account dividend payments and splits occurred in a particular day. Inspecting over a grid of values of b, W and ϕ, we have noted that the values of 0.9998, 22 and 1.125, respectively, yield values of q and H for $\{z_t\}$ that are in good agreement with a prior analysis of $\{r(t)\}$ which gave $q=1.48\pm0.02$ (using a simple t-Student distribution) and $q=1.51\pm0.02$ $\left(q'=1.47\pm0.003, v=0.92\pm0.008\right)$ [$\chi^2/n=0.00003$, $R^2=0.999$ and $P^*_{KS}=0.9276$](using the PDF of Eq. (9)) and persistence exponent $H=0.86\pm0.03$ (see Fig. 4). Comparing the numerical distribution of our model with the data we obtained $D_{KS}=0.014$ and a P^*_{KS} critical value equal to 0.991 from the two-sample Kolmogorov-Smirnov test [51], while the comparison between the distribution of the numerical procedure and the adjustment of the SP500 empirical distribution function yielded $P^*_{KS}=0.9998$. Once again we have tested other possible numerical adjustments and the only other relevant distribution was the stretched exponential with $v=1.3\pm0.02$ ($q'=1$) which has given a P^*_{KS} different from 1 ($P^*_{KS}=0.9999$), but a significantly larger value of χ^2 [$\chi^2/n=0.00009$, $R^2=0.9963$] (see Fig. 5).

It is worthy to be mentioned that all the three values of the parameters are plausible. First, within an application context, b is traditionally a value robustly greater than 0.9. Second, W is close to the number of business days in a month and last, but not least, ϕ is somewhat above the average level of the mean variance presented above. This provides us with a very interesting picture of the dynamics. Specifically, at a relevant approximation we can describe this particular system as monitoring the magnitude of its past fluctuations with a characteristic scale of a month, from which it computes the level of impact resulting in an excess of volatility. Actually one month moving averages are established indicators in quantitative analyses of financial markets. When the volatility in a period of the same order of magnitude of W surpasses the value $\phi/(1-b)$, then the system recalls previous periods of time, no matter how long they happened, in which a significant level of volatility excess occurred. Those periods are then averaged in order to determine the level of instantaneous volatility σ_t^2.
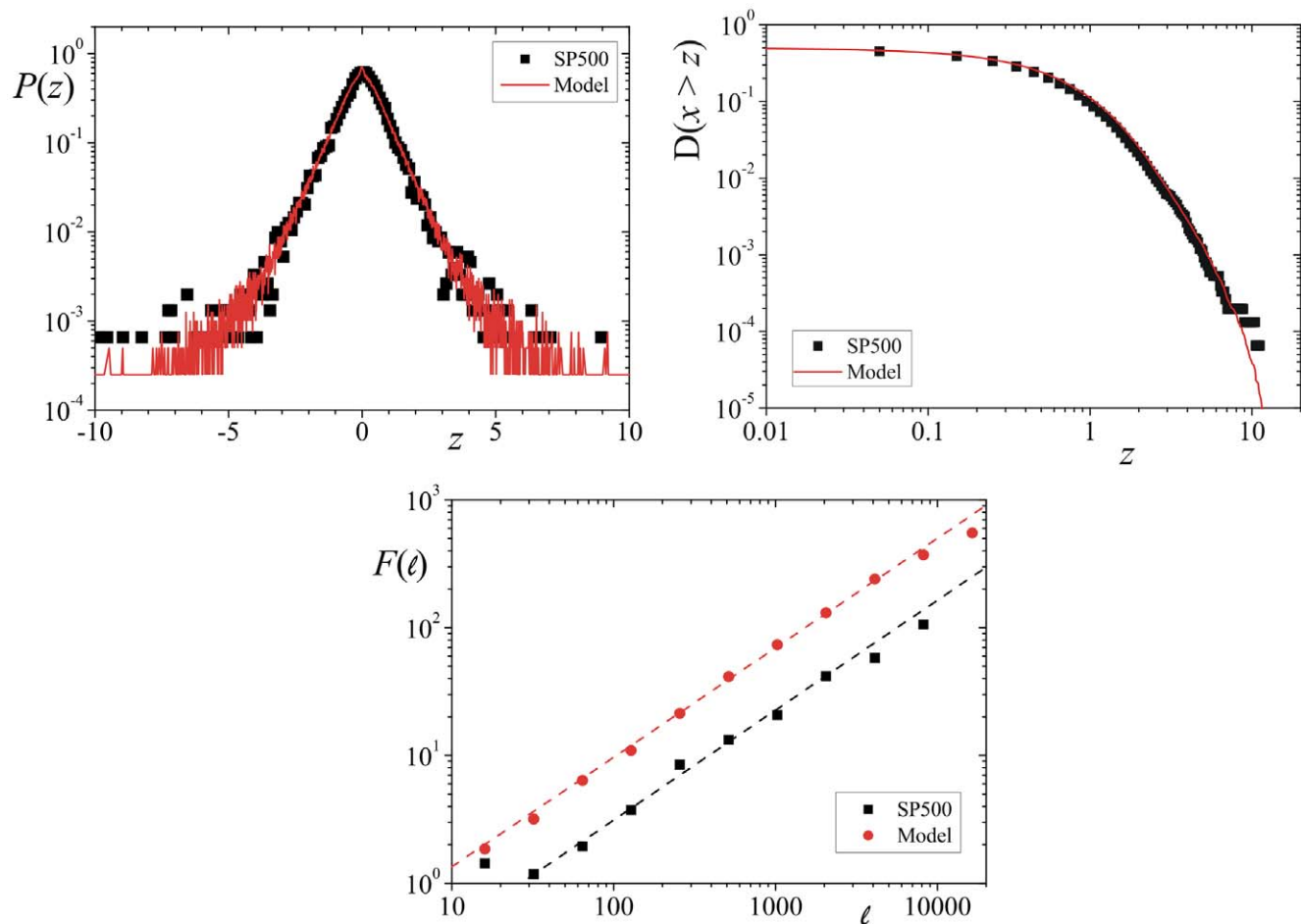
Discussion

We have studied a generalization of the well-known $ARCH$ process born in a financial context. Our proposal differs from other generalizations, since it adds to heteroscedastic dynamics the ability to reproduce systems where cognitive traits exist or systems showing typical cut-off limiting values. In the former case, when present circumstances are close to extreme and impacting events, the dynamics switches to the memory of abnormal events. By poring over the set of parameters of the problem, namely the impact of past values, b, the memory scale, W, and the volatility threshold, ϕ, we have verified that we are able to obtain times series showing fat tails for the probability density function and strong persistence for the magnitudes of the stochastic variable (directly related to the instantaneous volatility), as it happens in several processes studied within the context of complexity. In order to describe the usefulness of our model we have applied it to mimic the fluctuations of the stock index $SP500$, we verified that the best values reproducing the features of its time series are W close to one business month and ϕ greater that the mean variance of the process which is much larger than the normal level of volatility for which trading is not taken into account. Concerning the volatility, we have noticed that for the problems of interest (i.e., fat tails and

Figure 5. Probability density function $P(z)$ vs z. In the upper panels and on the left side we have $b=0.9998$, $W=22$ and $\phi=1.125$ (full line) $[q=1.49\pm0.01$ with $\chi^2/n=0.000025$ and $R^2=0.9984]$ and the $SP500$ daily log-index fluctuations (symbols) $[q=1.48\pm0.02$ with $\chi^2/n=0.00004$ and $R^2=0.996]$ in the log-linear scale and on the right side the complementary cumulative distribution function $D(x>z)$ vs z for case shown on the left. Lower panel: Fluctuation function $F(\ell)$ vs ℓ for the same parameters above $[H=0.85\pm0.02$, with $R=0.998]$ (red circles) and for the $SP500$ daily log-index fluctuations $[H=0.86\pm0.03$, with $R=0.997]$ (black squares) in a log-log scale.
doi:10.1371/journal.pone.0018149.g005

strong persistence), the distributions are very well described by a type-2 Gumbel distribution in large part of the domain, which explains the emergence of the tails.

Materials and Methods

Our results have been obtained from numerical simulation using code written in fortran language and run on the 64-bit ssolarII cluster (http://mesonpi.cat.cbpf.br/ssolar/).

Author Contributions

Conceived and designed the experiments: SMDQ EMFC FDN. Performed the experiments: SMDQ EMFC FDN. Analyzed the data: SMDQ EMFC FDN. Contributed reagents/materials/analysis tools: SMDQ EMFC FDN. Wrote the paper: SMDQ EMFC FDN.

References

1. Gell-Mann M (1996) The Quark and the Jaguar: Adventures in the Simple and the Complex. New York - NY: Abacus.
2. Flake GW (2000) The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems and Adaptation. Cambridge - MA: MIT Press.
3. Gell-Mann M, Tsallis C (2004) Nonextensive Entropy: Interdisciplinary Applications. New York - NY: Oxford University Press.
4. Skjeltorp AT, Vicsek T (2002) Complexity from Microscopic to Macroscopic Scales: Coherence and Large Deviations. Dordrecht: Kluwer Academic Publishers.
5. Bouchaud JP, Potters M (2000) Theory of Financial Risks: From Statistical Physics to Risk Management. Cambridge: Cambridge University Press.
6. Mantegna RN, Stanley HE (1999) An introduction to Econophysics: Correlations and Complexity in Finance. Cambridge: Cambridge University Press.
7. Voit J (2003) The Statistical Mechanics of Financial Markets. Berlin: Springer-Verlag.
8. Andersen TG, Bollerslev T, Diebold FX (2010) Parametric and nonparametric volatility measurement. In: Aït-Sahalia Y, ed. Handbook of Financial Econometrics. Amsterdam: Elsevier, volume 1, Tools and techniques. pp 67–102.
9. Khrennikov AY (2004) Information Dynamics in Cognitive, Psychological, Social, and Anomalous Phenomena. Dordrecht: Kluwer Academic Publishers.
10. Lee S, Ramenzoni VC, Holme P (2010) Emergence of collective memories. PLoS ONE 5: e12522.

11. Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50: 987.

12. Nobel Foundation. Available: http://nobelprize.org/nobel_prizes/economics/laureates/. Accessed 2011 Mar, 3.

13. Pobodnik B, Ivanov PCh, Lee Y, Cheesa A, Stanley HE (2000) Systems with correlations in the variance: Generating power law tails in probability distributions. Europhys Lett 50: 711.

14. Duarte Queirós SM, Tsallis C (2005) Bridging a paradigmatic financial model and nonextensive entropy. Europhys Lett 69: 893.

15. Thurner S, Farmer JD, Geanakoplos J (2009) Leverage causes fat tails and clustered volatility. arXiv:0908.1555 [q-fin.ST].

16. Donaldson PG, Kamstra M (1996) A new dividend forecasting procedure that rejects bubbles in asset prices: the case of 1929's stock crash. Rev Financ Stud 9: 333.

17. Hauser MA, Kunst RM (2001) Forecasting high-frequency financial data with the arfima-arch model. J Forecasting 20: 501.

18. Bollerslev T, Chou RY, Kroner KF (1992) Arch models in finance. J Econometrics 52: 5.

19. Ding Z, Granger CWJ, Engle RF (1983) A long memory property of stock market returns and a new model. J Emp Fin 1: 83.

20. Embrechts P, Kluppelberg C, Mikosch T (1997) Modelling Extremal Events for Insurance and Finance. Applications of Mathematics. Berlin: Springer-Verlag.

21. Gourieroux C, Montfort A (1996) Statistics and econometric models. Cambridge: Cambridge University Press.

22. Granger CWJ, Ding Z (1996) Modeling volatility persistence of speculative returns: a new approach. J Econometrics 73: 61.

23. Roman HE, Porto M (2008) Fractional brownian motion with stochastic variance: Modeling absolute returns in stock markets. Int J Mod Phys C 19: 1221.

24. Porto M, Roman HE (2001) Fractional brownian motion with stochastic variance: Modeling absolute returns in stock markets. Phys Rev E 63: 036128.

25. Duarte Queirós SM (2008) On a generalised model for time-dependent variance with long-term memory. EPL 80: 30005.

26. Dacorogna M, Gençay R, Müller U, Olsen R, Pictet O (2001) An Introduction to High-Frequency Finance. London: Academic Press.

27. Bouchaud JP (2002) An introduction to statistical finance. EPL 313: 238.

28. Hoskins B (2006) Predictability of Wheather and Climate, Cambridge: Cambridge University Press.

29. Podobnik B, Ivanov PCh, Biljakovic K, Horvatic D, Stanley HE, et al. (2006) Fractionally integrated process with power-law correlations in variables and magnitudes. Phys Rev E 72: 026121.

30. Campbell S, Diebold FX (2005) Weather forecasting for weather derivatives. J Am Stat Ass 100: 6.

31. Martin-Guerrero JD, Camps-Valls G, Soria-Olivas E, Serrano-Lopez EJ, Perez-Ruixo JJ, et al. (2003) Dosage individualization of erythropoietin using a profile-dependent support vector regression. IEEE Trans Biomed Eng 50: 1136.

32. Gronke P, Brehm J (2002) History, heterogeneity, and presidential approval: a modified arch approach. Elect Stud 21: 425.

33. Reynolds AM, Mordant N, Crawford AM, Bodenschatz E (2005) On the distribution of lagrangian accelerations in turbulent flows. New J Phys 7: 58.

34. Beck C (2007) Statistics of three-dimensional lagrangian turbulence. Phys Rev Lett 98: 064502.

35. Beck C (2001) Dynamical foundations of nonextensive statistical mechanics. Phys Rev Lett 87: 180601.

36. Zumbach G (2004) Volatility processes and volatility forecast with long memory. Quant Finance 4: 70.

37. Lux T, Marchesi M (1999) Scaling and criticality in a stochastic multi-agent model of a financial market. Nature 397: 498.

38. Giardina I, Bouchaud JP (2003) Bubbles, crashes and intermittency in agent based market models. Nature 31: 421.

39. Dose C, Porto M, Roman HE (2003) Autoregressive processes with anomalous scaling behavior: Applications to high-frequency variations of a stock market index. Phys Rev E 67: 067103.

40. Borland L (2004) A multi-time scale non-gaussian model of stock returns. arXiv:cond-mat/ 0412526.

41. Borland L (2010) Private communication.

42. Gradshteyn IS, Ryzhik IM (1965) Tables of integrals, series and products.

43. Hill BM (1975) A simple general approach to inference about the tail of a distribution. Ann Stat 3: 1163.

44. Clementi F, Di Matteo T, Gallegati M (2006) The power-law tail exponent of income distributions. Physica A 370: 49.

45. Tsallis C (1999) Nonextensive statistics: theoretical, experimental and computational evidences and connections. Braz J Phys 29: 1.

46. Beck C, Cohen EGD (2003) Superstatistics. Physica A 322: 267.

47. Queirós SMD, Tsallis C (2005) On the connection between financial processes with stochastic volatility and nonextensive statistical mechanics. Eur Phys J B 48: 139.

48. Queirós SMD (2008) On discrete stochastic processes with long-lasting time dependence in the variance. Eur Phys J B 66: 137.

49. Peng C-K, Buldyrev SV, Havlin S, Simons M, Stanley HE, et al. (1994) Mosaic organization of dna nucleotides. Phys Rev E 49: 1685.

50. Barunik J, Kristoufek L (2010) On hurst exponent estimation under heavy-tailed distributions. Physica A 389: 3844.

51. DeGroot MH (1991) Probability and Statistics. Reading - MA: Addison-Wesley.

52. The Wolfram functions site (2010) Meijer-g function. Available: http://functions.wolfram.com/. Accessed 2011 Mar 3.

53. Borland L, Bouchaud JP (2005) On a multi-timescale statistical feedback model for volatility fluctuations. arXiv:physics/0507073v1 [physics.soc-ph].