Autocorrelation and the sum of stochastic variables

Annibal Figueiredo [a] Iram Gleria [b] Raul Matsushita [c]
Sergio da Silva [d,1]

[a] *Department of Physics, University of Brasilia, Brazil*

[b] *Department of Physics, Federal University of Alagoas, Brazil*

[c] *Department of Statistics, University of Brasilia, Brazil*

[d] *Department of Economics, Federal University of Rio Grande do Sul, Brazil and National Council for Scientific and Technological Development, Brazil*

Abstract

We examine the role of nonlinear autocorrelations in the convergence to the Gaussian equilibrium and put forward an attempt to generalize the central limit theorem. Our results are illustrated with data coming from the British pound-US dollar rate.

Key words: Central Limit Theorem; Lévy distributions; Econophysics.
05.40.+j;02.50.-r

1 Introduction

We have previously suggested that scaling power laws associated with truncated Lévy flights (TLFs) [1] can be explained on the basis of particular features of autocorrelation in data [2]. Another interesting property of the TLF is a sluggish convergence to a Gaussian, observed in real data. Here we discuss the role of nonlinear autocorrelations in the convergence of the process and employ data from the British pound-US dollar rate to illustrate our point.

The paper is organized as follows. Section 2 presents our novel results. Section 3 exemplifies them, and Section 4 concludes.

[1] Corresponding author, e-mail:sergiodasilva@angelfire.com

2 Autocorrelation and the sum of stochastic variables

Consider the sum of random variables x_i, i.e.:

$$S_n \equiv \sum_{i=1}^{n} x_i$$

with zero mean for x_i. Lévy [3] shows that, for reduced variables $\bar{x} = x/\sqrt{\mu_2}$ (where μ_2 is the variance of x), the characteristic function (CF) $\psi(z)$ of a process with finite second moments can be written as $\psi(z) = e^{-z^2(1+w(z))/2}$, $w(0) = 0$. For the CF of x_i we can thus write

$$\psi_i(z) = e^{-\frac{\mu_{i2}}{2} z^2 (1+w_i(\mu_{i2}^{1/2} z))}$$

For S_n, the CF is

$$\Psi_n(z) = e^{-\frac{\nu_{n2}}{2} z^2 (1+\Omega_n(\nu_{n2}^{1/2} z))}$$

where ν_{n2} is the variance of S_n. The statistical moments of order p of x_i and S_n are $\mu_{ip} =< x_i^p >$ and $\nu_{np} =< S_n^p >$ respectively. Also consider that $\sigma_{np} = \sum_i \mu_{ip}$.

For independent variables it holds true that $\Psi_n(z) = \psi_1(z) \cdots \psi_n(z)$. But this does not hold for autocorrelated processes. Here we focus on a class of autocorrelated processes for which the CF of the sum variable S_n is such that

$$\Psi_n(z) = C_n(z)\psi_1(z) \cdots \psi_n(z) \tag{1}$$

where $C_n(z) = 1$ for an independent process.

Expanding the CF of x_i in series obtains

$$\psi_i(z) = 1 + \frac{\imath^2}{2!}\mu_{i2} z^2 + \frac{\imath^3}{3!}\mu_{i3} z^3 + \cdots \tag{2}$$

We assume that

$$C_n(z) = e^{-\frac{z^2}{2}(-2C_{n2}+W_n(z))} = 1 + C_{n2} z^2 + C_{n3} z^3 + \cdots \tag{3}$$

We can do the same for the CF of S_n. Performing the expansion of $\Psi_n(z)$ gives:

$$\Psi_n(z) = 1 + \frac{\imath^2}{2!}\nu_{n2} z^2 + \frac{\imath^3}{3!}\nu_{n3} z^3 + \cdots \tag{4}$$

Plugging (2), (3) and (4) in (1), and comparing equal order terms we obtain:

$$C_{n2} = -\frac{1}{2}(\nu_{n2} - \sigma_{n2}),$$

$$C_{n3} = -\frac{\imath}{2}(\nu_{n3} - \sigma_{n3}),$$

$$C_{n4} = \frac{1}{4!}(\nu_{n4} - \sigma_{n4}) - \frac{1}{2!2!}(\sigma_{n2}(\nu_{n2} - \sigma_{n2}) + \gamma_n) \tag{5}$$

where $\gamma_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mu_{i2}\mu_{j2}$. If we write $W_n(z) = \imath W_{n1} z + W_{n2} z^2 + O(z^3)$, from (3) and (5) one obtains

$$W_{n1} = \frac{1}{3}(\nu_{n3} - \sigma_{n3}), W_{n2} = \frac{1}{4}(\nu_{n2} - \sigma_{n2})^2 - 2C_{n4} \tag{6}$$

Plugging (5) and (6) back in the CF of S_n yields

$$\Psi_n(z) = e^{-\frac{1}{2}z^2\left(\nu_{n2} + \sum_{i=1}^{n} \mu_{i2} w_i(\mu_{i2}^{1/2} z) + W_n(z)\right)} \tag{7}$$

After writing the CF of the reduced variable as

$$\bar{\Psi}_n(z) = e^{-\frac{1}{2}z^2\left(1 + \Omega_n^{(1)}(z) + \Omega_n^{(2)}(z)\right)}$$

and reminding that

$$\bar{\Psi}_n(z) = \Psi_n\left(\frac{z}{\nu_{n2}^{1/2}}\right)$$

one has

$$\Omega_n^{(1)}(z) = \frac{1}{\nu_{n2}} \sum_{i=1}^{n} \mu_{i2} w_i \left(\left(\frac{\mu_{i2}}{\nu_{n2}}\right)^{1/2} z \right)$$

$$\Omega_n^{(2)}(z) = \frac{1}{\nu_{n2}} W_n \left(\frac{z}{\nu_{n2}^{1/2}} \right) \tag{8}$$

Function $\Omega_n^{(1)}$ matches the one for uncorrelated series, *i.e.* as $n \to \infty$ it approaches $w(0) = 0$ according to the central limit theorem (CLT). Term $\Omega_n^{(2)}(z)$ is related to the autocorrelations. It gives precisely the CF of the sum variable, which in turn can be used to obtain the probability distribution function (PDF) as $n \to \infty$. This generalizes the CLT for autocorrelated processes as in (1).

Now we relate $\Omega_n^{(2)}(z)$ to nonlinear correlation terms, which can be captured by

$$<p_1 p_2 \cdots p_k>_n = \sum_{i_1 \cdots i_k}^{n} (<x_{i_1}^{p_1} \cdots x_{i_k}^{p_k}> - <x_{i_1}^{p_1}> \cdots <x_{i_k}^{p_k}>) \tag{9}$$

where $p_1 p_2 \cdots p_k$ are positive integers, and $i_1 \neq i_2 \neq \cdots \neq i_k$. After writing $\Omega_n^{(2)} = \Omega_{n1}^{(2)} z\imath + \Omega_{n2}^{(2)} z^2$ it can be shown that

$$\Omega_{n1}^{(2)} = \frac{1}{3}\frac{\nu_{n3} - \sigma_{n3}}{\nu_{n2}^{3/2}} = \frac{1}{3}\frac{<111>_n + 3<12>_n}{\nu_{n2}^{3/2}} \tag{10}$$

$$\Omega_{n2}^{(2)} = \frac{1}{4}(1 - \frac{\sigma_{n2}^2}{\nu_{n2}^2}) - \frac{1}{12}\frac{\nu_{n4} - \sigma_{n4} - 6\gamma_n}{\nu_{n2}^2} = \tag{11}$$

$$\frac{1}{4}(1 - \frac{\sigma_{n2}^2}{\nu_{n2}^2}) - \frac{1}{12}\frac{<1111>_n + 6<112>_n + 4<13>_n + 3<22>_n}{\nu_{n2}^2}$$

where $\Omega_{n1}^{(2)}$ and $\Omega_{n2}^{(2)}$ are functions of third- and fourth-order autocorrelations respectively.

Due to the presence of nonlinear correlations, $\Omega_{n1}^{(2)}$ and/or $\Omega_{n2}^{(2)}$ may remain bounded above zero as n becomes larger. From these results it turns out that the limit distribution may not be a Gaussian. Furthermore, a suitable measure of the distance of a PDF to the Gaussian can be defined as follows. Given an arbitrary process with finite variance and moments $\mu_p =< x^p >$, let the CF of the reduced variable be written as

$$\bar{\psi}(z) = e^{-z^2(1+W(z))/2}$$

For a given δ, the distance between a given distribution f and the Gaussian can then be estimated as

$$D(f, Gauss) = \int_{-\delta}^{\delta} \sqrt{W_r(z)^2 + W_\imath(z)^2} dz \tag{12}$$

Expression $W(z) = W_r(z) + \imath W_\imath(z)$ can be expanded in series to give [2]

$$W_r(z) = -\frac{z^2}{12} K + O(z^4),$$

$$W_\imath(z) = \frac{z}{3} Sk + O(z^3)$$

where

$$K \equiv \frac{\mu_4}{{\mu_2}^2} - 3$$

is the kurtosis and

$$Sk \equiv \frac{\mu_3}{{\mu_2}^{3/2}}$$

is the skewness. Thus the leading terms in Ω_n are the kurtosis and skewness of the sum variable S_n. After remembering that such quantities are zero for a Gaussian, our results mean that the distance to the Gaussian is given by how far K and Sk are from zero, which sounds quite reasonable.

It has to be said that the data analysis is carried out by assuming a stationary process. This is particularly needed for the second moment to exist. Stationarity of the process is also implicitly presumed when the correlation structure between $x(t)$ and $x(t+k)$ is assumed to be the same, regardless of t and depending only on the lag k. And here it would really make sense average these values. Thus our methodology considers in particular stationary processes with long memory correlation structure. We leave for future research the issue of what is going on with conditionally heteroscedastic series that show volatility clusters.

3 Applications

Now we illustrate our approach with data coming from the daily variations of the British pound-US dollar rate. The data set contains 8033 data points, covering the time period from 1 April 1971 to 1 September 2003. We take returns Z rather than raw data as our stochastic variable, *i.e.*

$$Z_{\Delta t}(t) = Y(t + \Delta t) - Y(t)$$

where $Y(t)$ is a rate at day t. Note that $Z_{\Delta t}(t) \equiv S_n$ and $\Delta t \equiv n$.

Fig. 1 shows the curve of (12) for $\delta = 1$. It can be seen that the function is somewhat constrained to some real value which prevents termalization ($w(0) = 0$) to take place.

Fig. 2(a) presents the kurtosis and Fig. 2(b) the skewness. These are the leading terms in the expansion of $w(z)$. The curve of an IID process is shown for comparison. The skewness is clearly constrained to some real value; this in turn limits $\Omega^{(2)}_{n1}$ and then $\Omega^{(2)}_n = \Omega^{(2)}_{n1} z\imath + \Omega^{(2)}_{n2} z^2$. From (8) we conclude that the system cannot converge to the Gaussian in the time window from 1 until 500 trading days.

Figs. 3(a) and 3(b) present $\Omega^{(2)}_{n1}$ and $\Omega^{(2)}_{n2}$. From (8) and (11) it can be seen that when $\Omega^{(2)}_n \to \epsilon \neq 0$ the limit distribution is not Gaussian. From Figs.

3 one cannot say for sure that this is the case of our example, because we have stopped at $n = 500$. However, the fact that $\Omega_n^{(2)}$ is always different from zero in that time window do provide an explanation for the slow convergence in terms of nonlinear autocorrelations and the behavior of the kurtosis and skewness.

4 Conclusions

This paper examines the role of statistical autocorrelations in the convergence to the Gaussian equilibrium by focusing on the characteristic function; incidentally we generalize the central limit theorem. We explain the slow convergence in terms of both nonlinear autocorrelations and the behavior of the kurtosis and skewness. Our results are endorsed by exchange rate data from the pound-dollar daily returns.

5 Acknowledgements

We thank the valuable comments from an anonymous referee. Iram Gleria acknowledges support from FAPEAL.

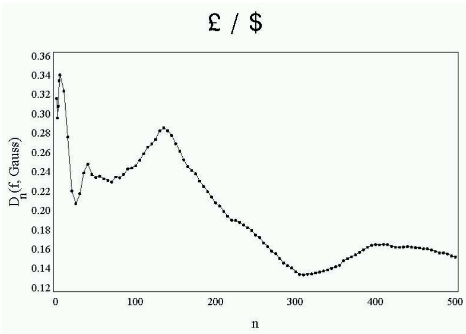

Fig. 1. (a) Distance to Gaussian distribution versus n (in trading days)

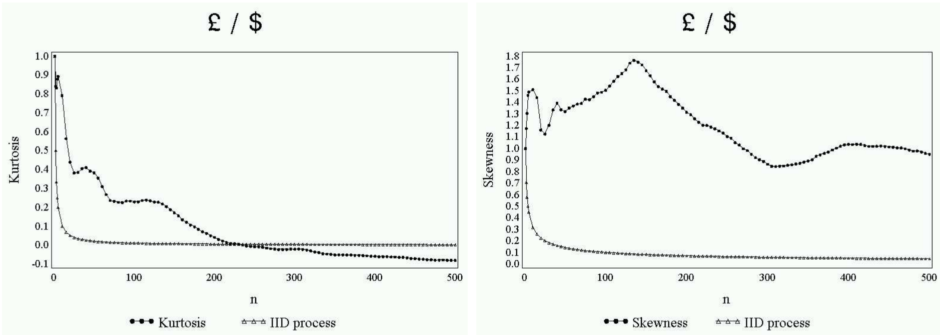

Fig. 2. (a) Behavior of the kurtosis for daily returns and an IID process. (b) Behavior of the skewness for daily returns and an IID process

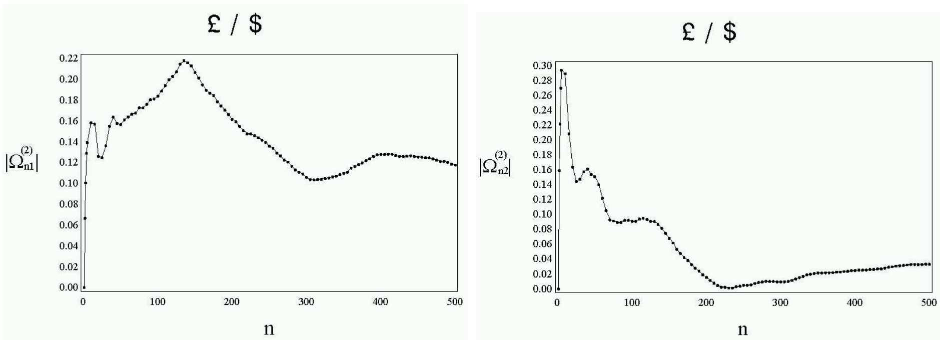

Fig. 3. (a) Norm of $\Omega_{n1}^{(2)}$ versus n. (b) Norm of $\Omega_{n2}^{(2)}$ versus n

References

[1] R. Mantegna, H. E. Stanley, *An Introduction to Econophysics, Correlations and Complexity in Finance*, (Cambridge University Press, Cambridge, 2000).

[2] A. Figueiredo, I. Gleria, R. Matsushita, S. Da Silva, Autocorrelation as a source of truncated Lévy flights in foreign exchange rates,*Physica A* **323** (2003) 601-625.

[3] D. Dugué, *Oeuvres de Paul Lévy, Vol. III: Eléments Aléatoires*, (Gauthiers-Villars, Paris, 1976).