Waiting-times and returns in high-frequency financial data: an empirical study

Marco Raberto [a], Enrico Scalas [b], Francesco Mainardi [c,1]

[a] *Dipartimento di Ingegneria Biofisica ed Elettronica, Università di Genova, via dell'Opera Pia 11a, I–16145 Genova, Italy*

[b] *Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale, via Cavour 84, I–15100 Alessandria, Italy*

[c] *Dipartimento di Fisica, Università di Bologna and INFN Sezione di Bologna, via Irnerio 46, I–40126 Bologna, Italy*

Abstract

In financial markets, not only prices and returns can be considered as random variables, but also the waiting time between two transactions varies randomly. In the following, we analyse the statistical properties of General Electric stock prices, traded at NYSE, in October 1999. These properties are critically revised in the framework of theoretical predictions based on a continuous-time random walk model.

Key words: PACS: 02.50.-r, 02.50.Ey, 02.50.Wp, 89.90.+n
Stochastic processes; Continuous-time random walk; Statistical finance; Econophysics; Autocorrelation function

1 Introduction

In financial markets, waiting times between two consecutive transactions vary in a stochastic fashion. In 1973, [1] Peter Clark wrote: "Instead of indexing [time] by the integers 0,1,2,..., the [price] process could be indexed by a set of numbers t_1, t_2, t_3, ..., where these numbers are themselves a realization of a stochastic process (with positive increments, so that $t_1 < t_2 < t_3 < \ldots$)."

From this point of view the continuous time random walk (CTRW) model of Montroll and Weiss [2] (see also Refs. [3–5]) can provide a phenomenological description of tick-by-tick dynamics in financial markets [6–8].

[1] Corresponding author. E-mail: `mainardi@bo.infn.it` URL: `www.fracalmo.org`

Actually in CTRWs, two random variables are used: jumps $\xi_n = x(t_{n+1}) - x(t_n)$ and waiting times $\tau_n = t_{n+1} - t_n$. In the financial interpretation of CTRWs, x represents a log-price and ξ a log-return [6–8] (see also [9]). The physicist can think of x as the position of a random walker performing discrete jumps in one dimension at randomly distributed instants. Based on [7] the evolution equation for $p(x,t)$, the probability of occurrence of the log-price x at time t, or of finding the random walker at position x at time instant t, can be written, assuming the initial condition $p(x,0) = \delta(x)$ (i.e. the walker is initially at the origin $x = 0$),

$$p(x,t) = \delta(x)\,\Psi(t) + \int_0^t \int_{-\infty}^{+\infty} \varphi(x-x', t-t')\, p(x',t')\, dx'\, dt' \,, \tag{1.1}$$

where $\Psi(t)$ is the *survival probability* and $\varphi(\xi,\tau)$, is the *joint probability density* of jumps $\xi_n = x(t_{n+1}) - x(t_n)$ and of waiting times $\tau_n = t_{n+1} - t_n$. Relevant quantities are the two (marginal) probability density functions (*pdf*'s) defined as $\lambda(\xi) := \int_0^\infty \varphi(\xi,\tau)\, d\tau$, $\psi(\tau) := \int_{-\infty}^{+\infty} \varphi(\xi,\tau)\, d\xi$, and called *jump pdf* and *waiting-time pdf*, respectively. If one assumes that the *jump pdf* $\lambda(\xi)$ is independent of the *waiting-time pdf* $\psi(\tau)$, we have the so-called "decoupling" which leads to the factorisation $\varphi(\xi,\tau) = \lambda(\xi)\,\psi(\tau)$.

The Eq. (1.1) is the most general *master equation* of the CTRW, usually derived in the Fourier-Laplace domain. The simplified form under the hypothesis of "decoupling" is reported in [7].

The probability that a given inter-step interval is greater or equal to τ is $\Psi(\tau)$, which is defined in terms of $\psi(\tau)$ by

$$\Psi(\tau) = \int_\tau^\infty \psi(t')\, dt' = 1 - \int_0^\tau \psi(t')\, dt' \,, \quad \psi(\tau) = -\frac{d}{d\tau}\Psi(\tau) \,. \tag{1.2}$$

We note that $\int_0^\tau \psi(t')\, dt'$ represents the probability that at least one step is taken at some instant in the interval $[0,\tau)$, hence $\Psi(\tau)$ is the probability that the diffusing quantity x does not change value during the time interval of duration τ after a jump. We also note, recalling that $t_0 = 0$, that $\Psi(t)$ is the *survival probability* until time instant t at the initial position $x_0 = 0$.

A relevant choice for the survival probability is given by the Mittag-Leffler function of order β $(0 < \beta < 1)$, which leads to a time-fractional master equation as shown in [7] (see also [10,11]). For reader's convenience hereafter we recall the main properties of this transcendental function useful for our purposes. From its definition valid for any $\beta > 0$:

$$\Psi(\tau) = E_\beta\left[-(\tau/\tau_0)^\beta\right] := \sum_{n=0}^\infty (-1)^n \frac{(\tau/\tau_0)^{\beta n}}{\Gamma(\beta n + 1)} \,, \quad \beta > 0 \,, \tag{1.3}$$

one recognises that the Mittag-Leffler function generalises the simple exponential function (recovered for $\beta = 1$) and, if $0 < \beta < 1$, it interpolates on the positive real axis a stretched exponential and a power law according to

$$E_\beta\left[-(\tau/\tau_0)^\beta\right] \sim \begin{cases} \exp\left[-(\tau/\tau_0)^\beta/\Gamma(1+\beta)\right], & \tau/\tau_0 \to 0^+, \\ (\tau/\tau_0)^{-\beta}/\Gamma(1-\beta), & \tau/\tau_0 \to \infty, \end{cases} \quad 0 < \beta < 1. \tag{1.4}$$

For more information on the Mittag-Leffler function see e.g. [12–14].

The purpose of this paper is to investigate some statistical properties of the random variables ξ and τ in financial markets. This study is limited to a specific equity of a given market in a definite period. Therefore caution is necessary and our results cannot be arbitrarily generalized. In particular, the reader will learn about General Electric stock prices, traded at NYSE, in October 1999. This preliminary presentation is part of a broader project aimed at studying the behaviour of all Dow-Jones-Industrial-Average stocks during that month.

2 Empirical analysis

In Fig. 1, a scatter plot is presented for waiting times τ_n as a function of the corresponding log-return ξ_n. By means of a contingency table analysis [15], we have studied the independence of the two stochastic variables. A direct inspection of Fig. 1 shows that for large values of log-returns waiting times tend to be shorter. This indicates a possible correlation. Actually, a hypothesis test has been performed on the empirical joint *pdf* $\varphi(\xi, \tau)$. According to the contingency table presented in Tab. 1, the two random variables cannot be considered independent. The null hypothesis of independence can be rejected with a significance level of 1%.

In Fig. 2, an estimate of the autocorrelation function for the absolute value of log-returns is plotted. We have used the following estimator [16]

$$C(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} (|\xi_{n+m}| - \overline{|\xi|})(|\xi_n| - \overline{|\xi|}), \tag{2.1}$$

where N is the total number of points ($N = 55559$) and $\overline{|\xi|} = \frac{1}{N}\sum_{n=0}^{N-1} |\xi_n|$. The inset shows the time series of the absolute values as a function of the tick n.

Due to scale persistence, the autocorrelation function follows a power-law decay with a slope of -0.76. The autocorrelation is over the noise level (conventionally $3/\sqrt{N}$) for a lag between 20 and 30 ticks, corresponding to an

		τ_n		
		$0 \div 10$	$10 \div 20$	> 20
ξ_n	< -0.002	25 (38.9)	21 (10.1)	9 (6.0)
	$-0.002 \div -0.001$	516 (613.6)	230 (159.5)	122 (94.9)
	$-0.001 \div 0$	6641 (7114.3)	2085 (1849.1)	1338 (1100.6)
	$0 \div 0.001$	31661 (31008.0)	7683 (8059.2)	4520 (4797.0)
	$0.001 \div 0.002$	398 (464.4)	179 (120.7)	80 (71.9)
	> 0.002	34 (36.1)	10 (9.4)	7 (5.6)

Table 1
Contingency table between log-returns ξ_n and waiting times τ_n. Every cell contains the frequency observed within the values considered and (in brackets) the theoretical frequency which can be computed under the null hypothesis of independence between ξ_n and τ_n.

average time of $250s$. Therefore, within that time scale, it is not safe to assume that the log-returns themselves, ξ_n, are independent variables. These are well-known stylised fact for tick-by-tick financial time series, see e.g. [17–19].

In Fig. 3, the autocorrelation function is shown for waiting-times τ_n. As above, the inset shows the time series itself. Waiting times between trades are inherently positive random variables. For the GE stock in October 1999, there is a marked seasonality of waiting times with a 1-day period (nearly $3,000$ trades). Inspection of the series shows that the trading activity is slower in the middle of the day. The seasonality is outlined by the periodic behaviour of the autocorrelation estimate, with periodicity above the conventional noise band.

In recent times, several efforts have been devoted to find a suitable *measure* of time, in order to discard similar seasonalities, see e.g. [20,21].

However, as shown in Fig. 4, the survival probability $\Psi(\tau)$ can be fitted by a *stretched exponential* function: $\exp\left[-(\tau/\tau_0)^\beta/\Gamma(1+\beta)\right]$, with $\tau_0 = 6.6s$ and $\beta = 0.7$. The reduced chi-square of the fit is 0.71.

In a previous work on bond futures [7], according to theoretical considerations on the properties of continuous-time random walks, we suggested the Mittag-Leffler function with a *power-law decay* as a suitable fit for the empirical survival probability. The above result does not contradict our previous findings. In fact, whereas for bonds futures we found waiting times greater than $10,000s$, here we have only waiting times smaller than $200s$, and the Mittag-Leffler function is well approximated by the *stretched exponential* as τ is small enough, see Eq. (1.4).

3 Summary

A preliminary study of General Electric high-frequency stock prices has been performed. Some statistical properties of the log-return and waiting-time random variables have been presented. This study was inspired by previous theoretical and empirical work, based on the phenomenological CTRW model of financial markets.

The main results are as follows: the two random variables cannot be considered independent from each other; the autocorrelation of log-returns absolute values exhibits a power-law decay and reaches the noise level after about 250 s; the autocorrelation of waiting times shows a 1-day periodicity, corresponding to the daily stock market activity.

References

[1] P.K. Clark, A subordinated stochastic process model with finite variance for speculative prices, Econometrica **41** (1973) 135-156.

[2] E.W. Montroll and G.H. Weiss, Random walks on lattices, II, J. Math. Phys. **6** (1965) 167–181.

[3] E.W. Montroll and B.J. West, On an enriched collection of stochastic processes, in: E.W. Montroll and J. Leibowitz (Eds.), Fluctuation Phenomena, North-Holland, Amsterdam, 1979, pp. 61-175.

[4] E.W. Montroll and M.F. Shlesinger, On the wonderful world of random walks, in: J. Leibowitz and E.W. Montroll (Eds.), Nonequilibrium Phenomena II: from Stochastics to Hydrodynamics, North-Holland, Amsterdam, 1984, pp. 1-121.

[5] G.H. Weiss, Aspects and Applications of Random Walks, North-Holland, Amsterdam, 1994.

[6] E. Scalas, R. Gorenflo and F. Mainardi, Fractional calculus and continuous-time finance, Physica A **284** (2000) 376-384.

[7] F. Mainardi, M. Raberto, R. Gorenflo, E. Scalas, Fractional calculus and continuous-time finance II: the waiting-time distribution, Physica A **287** (2000) 468-481.

[8] R. Gorenflo, F. Mainardi, E. Scalas and M. Raberto Fractional calculus and continuous-time finance III: the diffusion limit, in: M. Kohlmann and S. Tang (Eds.), Mathematical Finance, Birkhäuser, Basel, 2001, pp. 171-180.

[9] M. Parkinson, Option Pricing: The American Put, Journal of Business, **50** (1977) 21-36.

[10] R. Hilfer and L. Anton, Fractional master equations and fractal time random walks, Phys. Rev. E **51** (1995) R848–R851.

[11] A.I. Saichev and G.M. Zaslavsky, Fractional kinetic equations: solutions and applications, Chaos **7** (1997) 753–764.

[12] A. Erdélyi, W. Magnus, F. Oberhettinger, F.G. Tricomi, Higher Transcendental Functions, Bateman Project, Vol. 3, McGraw-Hill, New York, 1955. [Ch. 18, pp. 206-227.]

[13] R. Gorenflo and F. Mainardi, Fractional calculus, integral and differential equations of fractional order, in: A. Carpinteri and F. Mainardi (Eds.), Fractals and Fractional Calculus in Continuum Mechanics, Springer Verlag, Wien and New York, 1997, pp. 223–276. [down-loadable from: `http://www.fracalmo.org`]

[14] F. Mainardi and R. Gorenflo, On Mittag-Leffler type functions in fractional evolution processes, J. Comput. & Appl. Mathematics **118** (2000) 283-299.

[15] D.A. Berry and B.V. Linfgren, Statistics: Theory and Methods, Brooks/Cole Publishing, 1990.

[16] S.L. Marple, Digital Spectral Analysis with Applications, Prentice-Hall, Englewood Cliffs, N.J., 1987.

[17] R. Cont, Statistical properties of financial time series, Lectures presented at the Symposium on Mathematical Finance, Fudan University, Shangai, 10-24 August 1999. These lectures are down-loadable from
`http://www.eleves.ens.fr:8080/home/cont/papers.html`

[18] R.N. Mantegna and H.E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, Cambridge, UK, 2000.

[19] P. Gopikrishnan, V. Plerou, Y. Liu, L. Amaral, X. Gabaix, H.E. Stanley, Scaling and correlation in financial time series, Physica A **287** (2000) 362-373.

[20] G. Lefol and L. Mercier, Time deformation: definition and comparisons, Journal of Computational Intelligence in Finance **6** (1998) 19-33.

[21] M. Dacorogna, R. Genay, U.A. Mller, R.B. Olsen, O.V. Pictet, An Introduction to High-Frequency Finance, Academic Press, New-York, 2001.

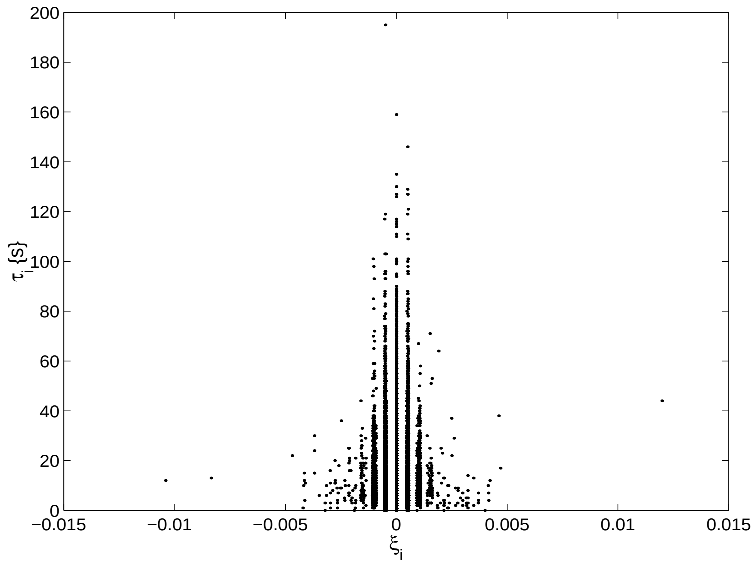

Fig. 1. Scatter plot of waiting times τ_n as a function of the corresponding log-returns ξ_n.

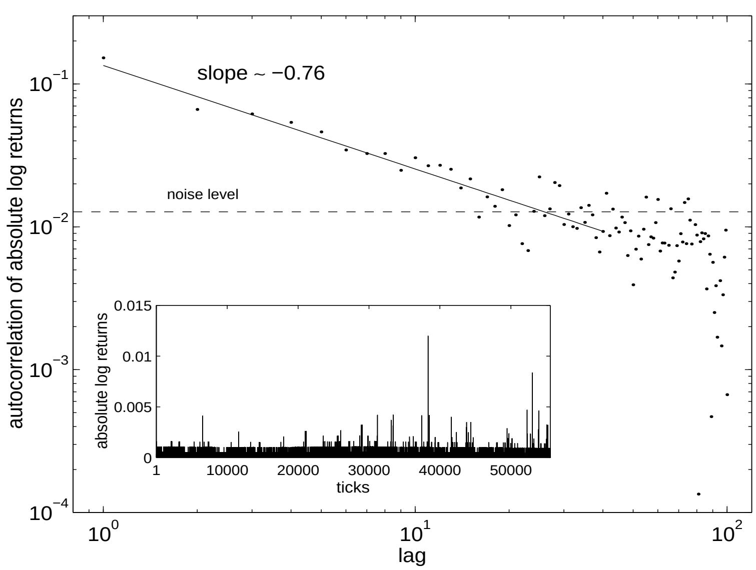

Fig. 2. Autocorrelation function for the absolute value of log-returns ξ_n. The inset shows the time series.

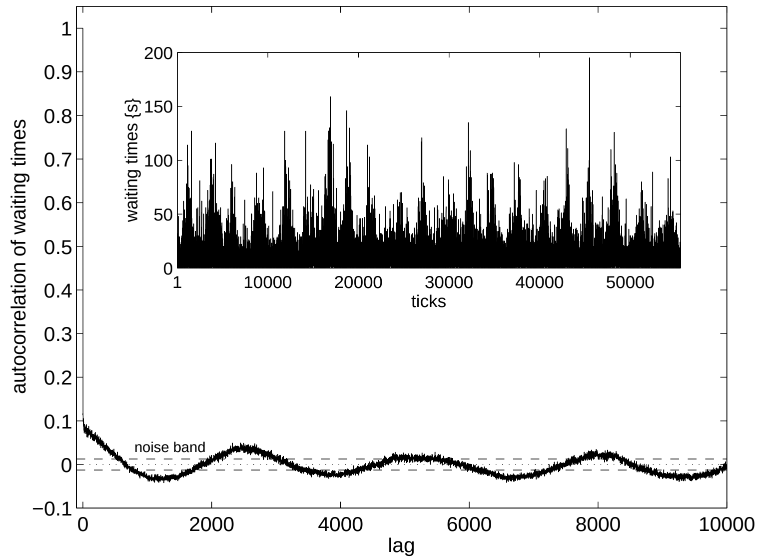

Fig. 3. Autocorrelation function for the waiting times τ_n. The inset shows the time series.

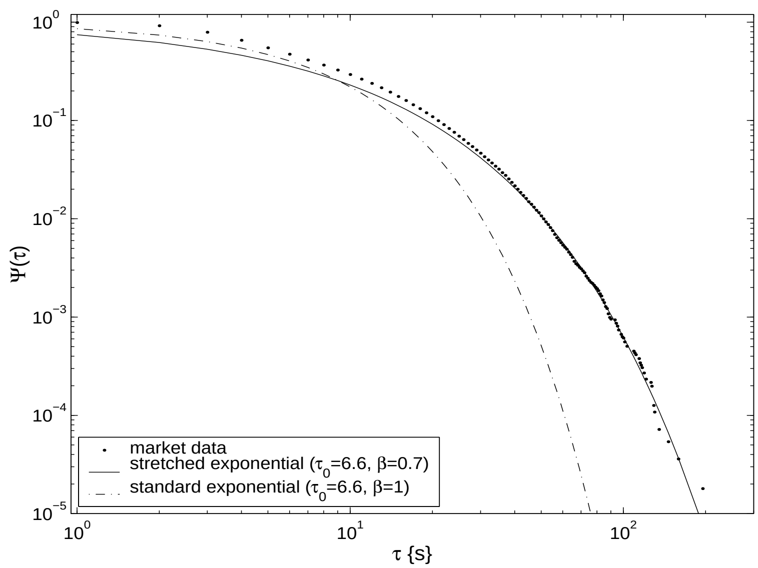

Fig. 4. Survival probability. The stretched exponential (solid line) is compared with the standard exponential (dash-dotted line).