Statistical analysis and agent-based microstructure modeling of high frequency financial trading

Linda Ponta, Member, IEEE, Enrico Scalas, Marco Raberto, and Silvano Cincotti*

Abstract—A simulation of high-frequency market data is performed with the Genoa Artificial Stock Market. Heterogeneous agents trade a risky asset in exchange for cash. Agents have zero intelligence and issue random limit or market orders depending on their budget constraints. The price is cleared by means of a limit order book. A renewal order-generation process is used having a waiting-time distribution between consecutive orders that follows a Weibull law, in line with previous studies. The simulation results show that this mechanism can reproduce fat-tailed distributions of returns without ad-hoc behavioral assumptions on agents. In the simulated trade process, when the order waiting-times are exponentially distributed, trade waiting times are exponentially distributed. However, if order waiting times follow a Weibull law, analogous results do not hold. These findings are interpreted in terms of a random thinning of the order renewal process. This behavior is compared with order and trade durations taken from real financial data.

Index Terms—High frequency financial time-series, Random thinning, Weibull distribution, Artificial Stock Market.

I. INTRODUCTION

In recent years, thanks to the availability of large databases of financial data, the statistical properties of high-frequency financial data and market microstructural properties have been studied by means of different tools, including phenomenological models of price dynamics and agent-based market simulations [1]–[18]. Various studies on high-frequency econometrics appeared in the literature including the autoregressive conditional duration models [19]–[23]. Among these approaches, agent-based based simulations [7], [8], [11]–[13] are particularly flexible as they allow the study of both the behaviour of agents and the influence of market structures in a well-controlled way. Since the early 1990s, artificial financial markets based on interacting agents have been developed. It is worth noting that besides some early Monte Carlo simulations (e.g., [24], [25]), microscopic simulations of financial markets initially aimed more to provide mechanisms for bubbles and crashes rather than to look at statistical features of the so generated time series. The first artificial market has been built at the Santa Fe Institute [26]–[28]. It is characterized by heterogeneous agents with limited rationality. While early attempts at microscopic simulations of financial markets appeared unable to account for the ubiquitous scaling laws of returns (and were, in fact, not devised to explain them), the recent models seem to be able to explain some of the statistical properties of financial data, but in most cases the attention is focused only on one stylized fact. Generally speaking, the objective of artificial markets is to reproduce the statistical features of the price process with minimal hypotheses about the intelligence of agents [29]. Several artificial markets populated with simple agents have been developed and have been able to reproduce some stylized facts, e.g., fat tails of returns and volatility autocorrelation [7], [8], [30]–[33]. For a detailed review on microscopic agent-based models of financial markets see [34], [35].

Stochastic models alternative to artificial markets have also been proposed, e.g., diffusive models, ARCH-GARCH models, stochastic volatility models, models based on fractional processes, models based on subordinate processes [36]–[42]. In particular, studies on stock-markets vulnerability by collective behaviour of large group of agents have been proposed. This led to consider collective behaviour that could reflect herding phenomena [36], [43], [44]. More recently, the role of heterogeneity, agents’ interactions and trade frictions on stylized facts of stock market returns have also been considered [45]. Here, the focus is on the influence of the double auction clearing mechanism where the price is fixed by the order book. An important empirical variable is the waiting time between two consecutive transactions [10], [46]. Empirically, in the market, during a trading day the activity is not constant [47] leading to fractal-time behavior [48], [49]. Due to the double auction mechanism, waiting times between two trades are themselves a stochastic variable [50]–[52]. They may also be correlated to returns [53] as well as to traded volumes. Indeed, trading via the order book is asynchronous and a transaction occurs only if a trader issues a market order. For liquid stocks, waiting times can vary in a range between some seconds and a few minutes, depending on the specific stock and on the market considered. In ref. [53], the reader can find a study on General Electric stocks traded in October 1999. Waiting times between consecutive prices exhibit 1-day periodicity, typical of variable intraday market activity. Moreover, the survival probability (the complementary cumulative distribution function) of waiting times is not exponential [54] but is well fitted by a Weibull function [19], [20].

In this paper, we simulate different distributions of waiting times between consecutive limit orders, namely the Weibull distribution and the power-law distribution. Orders are then selected by means of the limit order book mechanism implemented in the Genoa Artificial Stock Market (GASM) and

L. Ponta is with the Department of Physics, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy, e-mail: linda.ponta@polito.it.

E. Scalas is with Department of Science and Advanced Technology, University of East Piedmont, Via T. Michel 11, 15121 Alessandria, Italy, e-mail: enrico.scalas@mfm.unipmn.it and with BCAM - Basque Center for Applied Mathematics, Bizkaia Technology Park, Building 500, 48160 Derio, Spain, e-mail: email: escalas@bcamath.org

M. Raberto and S. Cincotti are with the Department of Biophysics and Electronics Engineering, University of Genova, Via Opera Pia 11a, 16145 Genova, Italy, e-mail: marco.raberto@unige.it, silvano.cincotti@unige.it. * To whom correspondence should be addressed.
described in Section II-B. The resulting distribution of waiting times between consecutive trades is then compared to a zero-order theory of order selection described in Section II-C. Section IV compares the simulation results with the empirical data extracted from the historical database of the London Stock Exchange. Finally, Section III is devoted to discussion and conclusions.

II. Model microstructure

In the implemented simulation, agents in the GASM trade one single stock in exchange for cash. They are liquidity traders and, therefore, the decision making process is nearly random and depends on the finite amount of cash plus stock available. At the beginning of the simulation, cash and stocks are uniformly distributed among agents.

A. Order generation

Trading is divided into $M$ daily sections. Each trading day is divided into $T$ elementary time steps of size one second. During the day, at given time steps $t_h$, a trader $k$ is randomly chosen to issue an order. Order waiting times $\tau_h^o = t_h - t_{h-1}$ are drawn according to a Weibull distribution. The Weibull probability density function is:

$$\varphi(\tau^o) = \frac{\beta}{\eta} \left( \frac{\tau^o}{\eta} \right)^{\beta-1} \exp\left[-\left(\frac{\tau^o}{\eta}\right)^{\beta}\right], \quad (1)$$

with $\tau^o > 0$, where $\eta$ is the scale parameter and $\beta$ is the shape parameter, also known as slope, as it is the slope of the regression line in a probability plot. The Weibull distribution reduces to the exponential distribution for $\beta = 1$. In these computational experiments we considered values of $\beta$ less than of equal to one. The order generation process is then described as a general renewal process where the waiting times between two consecutive orders, $\tau^o$, are independent and identically distributed (i.i.d.) random variables following (1). In the case $\beta = 1$, the order generation process is Poisson with an exponential waiting-time distribution. For further information on renewal processes the reader is referred to ref. [55].

B. Order selection and trading

A trader issues a buy or sell order with probability $1/2$. Let $a(t_{h-1})$ and $d(t_{h-1})$ be the values of the ask and bid prices stored in the book at time step $t_{h-1}$. In case the order issued at time step $t_h$ is a sell order, the limit price $s_k$ associated to the sell order is:

$$s_k(t_h) = n_k(t_h) \cdot a(t_{h-1}) \quad (2)$$

where $n_k(t_h)$ is a random draw by trader $k$ at time step $t_h$ from a Gaussian distribution with mean $\mu = 1$ and standard deviation $\sigma$. If $s_k(t_h) > d_k(t_{h-1})$ then the limit order is recorded in the book and no trade occurs, else the order becomes a market order and a transaction takes place at the price $S(t_h) = d(t_{h-1})$. In the latter case, the sell order is partially or totally fulfilled and the bid price is updated. The quantity of stock offered for sale is a random fraction of the quantity owned by the trader. In case the order is a buy order, the limit price $b_k(t_h)$ is now:

$$b_k(t_h) = n_k(t_h) \cdot d(t_{h-1}), \quad (3)$$

where $n_k(t_h)$ is determined as above. If $b_k(t_h) < a_k(t_{h-1})$ then the limit order is recorded in the book and no trade occurs, else the order becomes a market order and a transaction takes place at the price $S(t_h) = a(t_{h-1})$. The quantity of stock ordered depends on the cash of trader $k$ and on the value of $b_k(t_h)$. In this framework, agents compete for liquidity. If a buy order is issued by an agent, its benchmark is the best limit buy order given the bid price. As $\mu = 1$, for half of the times, the agent offers a more competitive buy order (if $b_k(t_h) > d(t_{h-1})$), that can result in a trade if $b_k(t_h) \geq a(t_{h-1})$. The same is valid for sell limit orders.

C. Random thinning for order selection and trading

As a zero-order model of order selection and trading, let us consider the random thinning [55] of the order generation process. This has been studied by Gnedenko and Kovalenko [56]. In order to define the thinning procedure, one first defines the epochs of events (orders) $t_n^o$ as

$$t_n^o = \sum_{i=1}^{n} \tau_i^o. \quad (4)$$

Then, the sequence $t_1^o, t_2^o, \ldots, t_n^o, \ldots$ is decimated according to this rule. Every epoch is independently kept with probability $q$ or deleted with probability $p = 1 - q$ with $0 < q < 1$. In order to compute the probability density function $(T_q \varphi)(\tau)$, the probability density function $f_k(t)$ of the sum of $k$ waiting times is needed. As waiting times are i.i.d. variables, $f_k(t)$ is given by the $k$-fold convolution of $\varphi$:

$$f_k(t) = \varphi(t), \quad f_k(t) = \int_0^t f_{k-1}(t - t') \varphi(t') dt'. \quad (5)$$

$(T_q \varphi)(\tau)$ can be obtained by purely probabilistic arguments, by noting that, after a kept event, the next one of the original process is kept with probability $q$ but dropped in favor of the second next with probability $pq$ and, in general, $n - 1$ events are dropped in favor of the $n$-th next with probability $p^{n-1}q$. Therefore one has:

$$(T_q \varphi)(\tau) = \sum_{n=1}^{\infty} qp^{n-1} f_n(\tau). \quad (6)$$

Let $\hat{f}_n(s) = \int_0^\infty e^{-st} f_n(t) dt$ be the Laplace transform of $f_n(t)$. From the behavior of the Laplace transform of a convolution, it turns out that the Laplace transform of eq. (6) is:

$$(T_q \hat{\varphi})(s) = \sum_{n=1}^{\infty} qp^{n-1}[\hat{\varphi}(s)]^n = \frac{q \hat{\varphi}(s)}{1 - (1-q) \hat{\varphi}(s)}, \quad (7)$$

from which, in principle, we can reconstruct by Laplace inversion the probability density function of the thinned process. Eq. (6) or eq. (7) can be used to estimate the density of waiting times between two consecutive transaction from the knowledge of the order waiting-time probability density.
function. Alternatively, a Monte Carlo simulation can be used, in which the thinning procedure is preformed directly on a pseudo-random sequence of waiting times. Random thinning for order selection and trading is a rough approximation as it does not take into account many features which are present in a market, including price and volume feedback and partial order fulfillment. However the distributions obtained by random thinning can be easily generated and compared with those obtained with GASM by the procedure described in Section II-B.

III. SIMULATION RESULTS

The simulations are performed with the following parameters. The number of daily sections $M$ is set equal to 50. The length of the daily sections is $T = 25,200$ s (corresponding to 7 hours of trading activity). In Fig. 1, data are presented for Weibull-distributed orders with $\beta = 1$ (exponential case). In Fig. 2, the case $\beta = 0.4$ is discussed. The average waiting-time $\langle \tau^o \rangle$ between orders is set to 20 s for every simulation. The scale factor $\eta$ is related to $\langle \tau^o \rangle$ according to

$$\eta = \frac{\langle \tau^o \rangle}{\Gamma(1/\beta + 1)},$$

(8)

where $\Gamma(\cdot)$ is Euler’s Gamma function. The survival probability $P_\tau(\tau)$ corresponding to the Weibull density (1) is

$$P_\tau(\tau) = \exp(-\tau/\eta)^\beta.$$

(9)

The lifespan of orders is 600 seconds, a time much larger than $\langle \tau^o \rangle$. Sell and buy limit prices are computed following (2) and (3), respectively. The random numbers $n_i(t_k)$ are drawn from a normal distribution with parameters $\mu = 1$ and $\sigma = 0.005$. The number of agents is 10,000. The initial stock price is 100.00 units of cash, say Euros, and each trader owns an equal amount of cash and shares: 100,000 Euros and 1,000 shares. These simulations produce realistic intraday price paths [8]. In Fig. 1, the survival probability distribution of order waiting times is compared with that for trade waiting times with $\beta = 1$. This case corresponds to exponentially distributed order waiting times. As a consequence of the GASM order selection procedure, the waiting time between trades, $\langle \tau^t \rangle$, is still exponentially distributed, with a larger average waiting time. The hollow circles in Fig. 1 correspond to a Monte Carlo simulation of random thinning with a probability $q = \langle \tau^o \rangle/\langle \tau^t \rangle = 0.36$. The agreement between the GASM order selection and the random thinning procedure is good.

In Fig. 2, the survival probability distribution of order waiting times is compared with that for trade waiting times with $\beta = 0.4$. This case corresponds to Weibull-distributed order waiting times. As a consequence of the GASM order selection procedure, the waiting time between trades, $\langle \tau^t \rangle$, no longer follows the Weibull distribution. In Fig. 2, the dashed line is the Weibull fit of the trade waiting-time survival function and a Kolmogorov-Smirnov test rejects the null hypothesis of Weibull-distributed trade waiting times at the 5% significance level. The hollow circles in Fig. 2 correspond to a Monte Carlo simulation of random thinning with a probability $q = \langle \tau^o \rangle/\langle \tau^t \rangle = 0.36$. Again, the agreement between the GASM order selection and the random thinning procedure is good.

IV. EMPIRICAL ANALYSIS

In this section, simulation results are compared to the behaviour of real data for the sake of completeness. The waiting-time empirical data have been extracted from the historical database of the London Stock Exchange where orders and quotes are stored for the electronic market; these data are a significant fraction, but do not include all the orders and quotes. The data set analyzed consists of waiting times between orders and trades for both Glaxo Smith Kline (GSK) and Vodafone (VOD) stocks traded in the following months: March, May, and October 2002. Both limit and market orders have been included. The use of one-month high-frequency data is a trade-off between the necessity of managing enough data for significant statistical analysis and, on the other hand, the goal of minimizing the effect of external economic fluctuations. Figs. 3 and 4 show the waiting-time survival functions for the orders (blue dots), trades (blue
crosses) and the results of the random thinning (red hollow circles) of the GSK and VOD stocks respectively. The blue line represents the Weibull fit of orders, the blue dashed line the one of trades and the red point line the one of the thinning results. The empirical analysis summarized in Figs. 3 and 4 shows that the random thinning of orders approximately reproduces the statistical behavior of trade duration. For other data, [52], [57], the waiting time of trade duration follows the Weibull distribution. However, in our case, Weibull distribution fits, performed with the moment methods, are presented in Figs. 3 and 4. The results show that neither order nor trade durations follow the Weibull distribution. These findings are also corroborated by Kolmogorov-Smirnov test rejecting the null hypothesis of Weibull distribution data.

V. DISCUSSION AND CONCLUSIONS

The simulation results described in Section III can be interpreted as follows. When the waiting-time distribution between orders is exponential, then the GASM order selection procedure described is Section II-B leads to exponentially distributed waiting times between consecutive traders. When the order waiting times follow a Weibull renewal process with $0 < \beta < 1$, then the trade waiting-time distribution is no longer ruled by the Weibull law. However, in both cases, in regard to waiting times, the outcome of the order selection process is well-mimicked by a simple random thinning with probability $q$ given by the ratio between the average order waiting time and the average trade waiting time: $q = \langle \tau^o \rangle / \langle \tau^t \rangle$. In other words, the GASM selection process of the order book is equivalent to a random thinning for the simulation parameters investigated.

In other words, the GASM selection process of the order book is equivalent to a random thinning for the simulation parameters investigated. This random thinning procedure seems to work also for empirical data, meaning that it is able to predict the unconditional distribution of trade durations from the knowledge of the unconditional distribution of order durations. It could be argued that there is no strong reason for independent market investors to place buy and sell orders in a time-correlated way. This argument would lead one to expect a Poisson process for orders. Therefore, if price formation were a simple thinning of the bid-ask process, then exponential waiting times should be expected between consecutive trades as well [55]. Eventually, even if empirical analysis should show that time correlations are already present at the order level, it would be interesting to understand why they are there. In other words, the empirical results on the survival probability set limits on statistical market models for price formation. A possibly correlated result has been recently obtained by Fabrizio Lillo and Doyne Farmer, who find that the signs of orders in the London Stock Exchange obey a long-memory process [58]–[60]. However, non-exponential unconditional survival probability can also be explained by a mixture of exponentials due to variable activity during a trading day. In this case, one has

$$P_\geq(\tau) = \sum_{i=1}^{N} a_i e^{-\mu_i \tau}, \quad (10)$$

where $a_i$ are suitable weights whose sum $\sum_{i=1}^{N} a_i$ must be 1, to fulfill the condition $P_\geq(0) = 1$ [10]. Further empirical studies on market microstructure will be necessary to clarify these points.

ACKNOWLEDGMENT

This work has been partially supported by the University of Genoa and by the Italian Ministry of Education, University and Research (MIUR) under grants COFIN2007. The Istituto Superiore Mario Boella is gratefully acknowledged for financial support. The LSE duration data were made available by the Italian project "Dinamica di altissima frequenza dei mercati..."
finanziari”. E.S. wishes to thank Alessandra Tedeschi for her skillful work on these data sets.

REFERENCES


Fig. 4. Survival probability distribution of order waiting times (dots) and of trade waiting times (crosses) for the VOD stock of LSE for March 2002 (a), May 2002, (b) and October 2002 (c). The hollow circles represent the results of random thinning with probability equal to $q$ equal to the ratio between the length of the trade and the orders series. The three lines represent the corresponding Weibull fits.


