Communications in Applied and Industrial Mathematics, DOI: 10.1685/journal.caim.535
ISSN 2038-0909, e-535

Large-scale simulations of synthetic markets

Luca Gerardo-Giorda[1], Guido Germano[2,3], Enrico Scalas[1,4]

[1] *BCAM — Basque Center for Applied Mathematics*
Bilbao, Basque Country, Spain
lgerardo@bcamath.org

[2] *Department of Computer Science*
University College London, UK
g.germano@ucl.ac.uk

[3] *Systemic Risk Centre,*
London School of Economics, UK
g.germano@lse.ac.uk

[4] *School of Mathematical and Physical Sciences*
Department of Mathematics
University of Sussex, UK
e.scalas@sussex.ac.uk

Communicated by Gianni Pagnini

Abstract

High-frequency trading has been experiencing an increase of interest both for practical purposes within financial institutions and within academic research; recently, the UK Government Office for Science reviewed the state of the art and gave an outlook analysis. Therefore, models for tick-by-tick financial time series are becoming more and more important. Together with high-frequency trading comes the need for fast simulations of full synthetic markets for several purposes including scenario analyses for risk evaluation. These simulations are very suitable to be run on massively parallel architectures. Aside more traditional large-scale parallel computers, high-end personal computers equipped with several multi-core CPUs and general-purpose GPU programming are gaining importance as cheap and easily available alternatives. A further option are FPGAs. In all cases, development can be done in a unified framework with standard C or C++ code and calls to appropriate libraries like MPI (for CPUs) or CUDA for (GPGPUs). Here we present such a prototype simulation of a synthetic regulated equity market. The basic ingredients to build a synthetic share are two sequences of random variables, one for the inter-trade durations and one for the tick-by-tick logarithmic returns. Our extensive simulations are based on several distributional choices for the above random variables, including Mittag-Leffler distributed inter-trade durations and alpha-stable tick-by-tick logarithmic returns.

Keywords: `synthetic markets, large scale simulation, heteroskedasticity`

AMS subject classification: `60G50, 60F17, 33E12, 26A33`

Received on 04 01, 2014. Accepted on 06 30, 2015. Published on 08 10, 2015.

1. Introduction

1.1. *General considerations*

High-frequency trading is a field that experienced an explosion of interest within financial institutions, for practical purposes, and within academic scholars, as well. The interested reader can consult a recent report by Furse et al. [3] for a review of the state of the art and an outlook analysis. Therefore, models for tick-by-tick financial fluctuations [2] are becoming more and more important. A review of this literature can be found in a recent book [6]. Along with high-frequency trading comes the need for fast simulations of full synthetic markets for several purposes including scenario analyses for risk evaluation.

Some years ago, two authors of this paper, G.G. and E.S., proposed the use of distributed computing for large-scale tick-by-tick synthetic market simulation [5] also by exploiting the embarrassingly parallel character of Monte Carlo simulations. Nowadays, rather new techniques and hardware, such as graphics processsing units (GPUs) [10] and field programmable gate arrays (FPGAs) [7], are available to assist decision making in high-frequency and intra-day trading. Moreover, with the development of new standard libraries for GPUs and FPGAs, it becomes no longer necessary to develop programs in dedicated programming languages and one can use code written in `C` or `C++` with appropriate library calls. Suitable interpreters or compilers port the software to the chosen hardware devices for fast and reliable execution. In this paper, we show the simulation of synthetic *regulated equity markets*, and in the following section we describe their modeling.

1.2. *Modeling a synthetic market*

1.2.1. *Single equity*

The behavior of a single equity in time is a combination of the time lapse between two subsequent trades and the price variation at each trade. This is due to the double-auction mechanism implemented in equity markets [11]. A synthetic share dynamics can therefore be modeled by the combined effect of two sequences of random variables defined as follows.

Definition 1.1. (Inter-trade durations) The inter-trade durations are defined as a sequence $\{J_i\}_{i=1}^{\infty}$ of positive real random variables.

Definition 1.2. (Tick-by-tick logarithmic returns) The tick-by-tick logarithmic returns are defined as a sequence $\{Y_i\}_{i=1}^{\infty}$ of real random variables.

Remark 1.1. (Dependence) At this stage, no assumption is made on the dependence structure of the random variables defined above.

DOI: 10.1685/journal.caim.535

Remark 1.2. (Continuous double auction) The majority of regulated equity markets use the continuous double auction to determine trade prices. This leads to asyncronous trading epochs; in other words not only the returns are random variables but also inter-trade intervals.

With these two basic ingredients, we can define trading epochs

Definition 1.3. (Trading epochs) Let $\{J_i\}_{i=1}^{\infty}$ be inter-trade durations, then the sequence of trading epochs or events $\{T_n\}_{n=1}^{\infty}$ is defined as follows

$$T_n = \sum_{i=1}^{n} J_i, \tag{1}$$

and the counting process is defined as

Definition 1.4. (Counting process) Let $\{T_n\}_{n=1}^{\infty}$ be a sequence of trading epochs, then the process $N(t)$ counting the number of events up to time t is given by

$$N(t) = \max\{n : T_n \leq t\}. \tag{2}$$

Now, one can define the logarithmic-price process as

Definition 1.5. (Logarithmic-price process) Let $\{Y_i\}_{i=1}^{\infty}$ be a sequence of tick-by-tick logarithmic returns and let $N(t)$ be the counting process corresponding to the trading epochs $\{T_n\}_{n=1}^{\infty}$, then the logarithmic-price process $X(t)$ is

$$X(t) = \sum_{i=1}^{N(t)} Y_i = \sum_{i=1}^{\infty} Y_i \mathbf{1}_{\{T_i \leq t\}}. \tag{3}$$

We are ready to define a synthetic share process or price process

Definition 1.6. (Synthetic share) Let $X(t)$ be a logarithmic-price process and let S_0 be the opening price (at epoch $T_0 = 0$) of the share (a given positive real variable), then the synthetic share process or price process is

$$S(t) = S_0 \mathrm{e}^{X(t)}. \tag{4}$$

Remark 1.3. (Opening price) In many regulated equity markets, the opening price or open is fixed by an appropriate opening auction which does not use the continuous double auction mechanism.

1.2.2. *Whole market*

The above description of the dynamics of a synthetic share process can be easily generalized to a procedure describing the behaviour of an arbitrary number of shares M, representing the number of traded assets in an equity market.

Definition 1.7. (Synthetic market) Let $\{S^{(k)}(t)\}_{k=1}^{M}$ be a collection of synthetic share processes defined as above. We say that this collection represents a synthetic market.

Remark 1.4. (Again on dependence) No assumption has been made on the dependence of the price processes making up the synthetic market. In general, these processes will not be independent, but their dependence structure may evolve in time [9].

Remark 1.5. (Specification of processes) In order to derive theorems and make calculations on synthetic markets, it is necessary to specify a market process. For instance, we have studied the share process when the inter-trade durations and the tick-by tick logarithmic prices are independent and identically distributed and mutually independent. In this simple case, the price process is semi-Markov and many properties can be derived [1,4,12,14].

1.2.3. *Synthetic market simulation and assessment*

Based on the construction outlined above, we use the following 4-steps procedure to build effective synthetic markets for high-frequency trading applications.

1. **Selection**: Select an intra-day market model with full specification of the market process.
2. **Fitting**: Fit the parameters of the market process with the available historical data.
3. **Simulation**: Run Monte Carlo simulations of the fitted synthetic market.
4. **Assessment of quantities of interest**: Compute the quantities of interest out of the Monte Carlo simulations.

Given the non-stationary nature of financial markets, this procedure does not protect against *unexpected* events, but it is the best thing one can do, also considering that the second step can be made adaptive and the fitting procedure can be constantly updated during continuous trading. However, in this paper, we do not discuss this interesting and difficult issue, nor are we presenting rules or advices for the first step (the selection of the market

DOI: 10.1685/journal.caim.535

model). Rather, we are interested in showing that the third and fourth steps can be easily performed. This will be the subject of the next section.

2. Simulations

2.1. *General framework*

We consider a synthetic market consisting of $N = 2\,000$ different shares. For the sake of simplicity, here we consider mutually independent couples (J_i, Y_i) as well as mutually independent log-prices. Dependences among stock prices can be conveniently introduced using factor models [8]. We simulate a variety of scenarios, ruled by different choices for the inter-trade durations and logarithmic returns. In particular, for any share (labeled with $s = 1, \ldots, N$), the intra-trade duration epochs $J_{i,s}$ are draws of a random variable J following one of the following distributions.

1. *Mittag-Leffler* distribution, characterized by the survival function

$$\mathbb{P}(J > u) = E_\beta(-u^\beta), \tag{5}$$

where $E_\beta(z)$ is the one-parameter Mittag-Leffler function

$$E_\beta(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\beta n + 1)}, \qquad z \in \mathbb{C}. \tag{6}$$

2. *ACD(p,q) (Autoregressive Conditional Duration)* distribution. The durations are given by

$$J_t = \Theta_t Z_t, \tag{7}$$

where Z_t are positive, independent and identically distributed random variables, with $\mathbb{E}(Z_t) = 1$. The time series Θ_t is given by

$$\Theta_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j J_{t-j} + \sum_{i=1}^{p} \beta_i \, \Theta_{t-i}, \tag{8}$$

where $\alpha_0 > 0$, $\alpha_j \geq 0$ $(j = 1, \ldots, q)$, $\beta_i \geq 0$ $(i = 1, \ldots, p)$.

On the other hand, the tick-by-tick logarithmic returns Y_t are given by a time series modeled as follows.

1. *Independent and identically distributed Lévy α-stable random variables.* We let $Y_t \sim L_\alpha$, where L_α is a random variable with probability density function given by

$$f_{L_\alpha}(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-|\kappa|^\alpha} \, e^{-i\kappa x} \, d\kappa. \tag{9}$$

Remark 2.1. Processes with α-stable i.i.d. increments that are subordinated to the limit of a Mittag-Leffler process have a one-point probability density function converging to the solution of the fractional diffusion equation.

2. *ARCH(q) (Auto-Regressive Conditional Heteroskedasticity)*: we let $Y_t \sim \epsilon_t$, where the last term is split into a stochastic term z_t and a time-dependent standard deviation σ_t. The *ARCH* model of order q [2] reads

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \sum_{j=1}^{q} \alpha_j (\epsilon_{t-j})^2 \\ \epsilon_t &= \sigma_t z_t, \end{aligned} \tag{10}$$

where $\alpha_j > 0$ $(j \geq 0)$, while z_t is a standard white noise $z_t \sim N(0,1)$.
3. *GARCH(p,q) (Generalized Auto-Regressive Conditional Heteroskedasticity)*: we let again $Y_t \sim \epsilon_t$, where the last term is split into a stochastic term z_t and a time-dependent standard deviation σ_t. The *GARCH* model of order (p,q) is obtained assuming an autoregressive moving average model (*ARMA*) for the variance of the error, and reads [2]

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \sum_{j=1}^{q} \alpha_j (\epsilon_{t-j})^2 + \sum_{i=1}^{p} \beta_i \, \sigma_{t-i}^2 \\ \epsilon_t &= \sigma_t z_t, \end{aligned} \tag{11}$$

where $\alpha_j > 0$ $(j = 0, \ldots, q)$, $\beta_i > 0$ $(i = 1, \ldots, p)$, and again z_t is a standard white noise $z_t \sim N(0,1)$.

Remark 2.2. Auto-Regressive Conditional Heteroskedasticity (ARCH) models are used whenever there is reason to believe that, at any point in a series, the terms will have a characteristic size or variance. For this reason, such models are very popular in modeling financial time series that exhibit time-varying volatility clustering.

2.2. *Setup of test scenarios*

In this section we present four different scenarios simulated from different combinations of the methods presented earlier, as reported in Table 1, where we summarize the choices of the inter-trade duration distribution and the logarithmic returns.

For the inter-trade duration we use both a Mittag-Leffler and a *ACD(1,1)* model, whose stability is guaranteed provided $\alpha_1 + \beta_1 < 1$.

DOI: 10.1685/journal.caim.535

The tick-by-tick logarithmic returns are drawn from a i.i.d. Lévy α-stable process (with $\alpha \in (1,2)$ [13,15]), an *ARCH(1)*, and a *GARCH(1,1)* model. The stability of the latter model is guaranteed provided $\alpha_1 + \beta_1 < 1$.

2.3. *Simulations and results*

We performed a Monte Carlo simulation with 10 000 realisations of the whole market, where, for each realisation, the parameters of the runs are generated by a random algorithm, with 2 000 draws from the distributions summarized in Table 2. Finally, to simulate the individual behavior, we associate each of the 2 000 shares in the synthetic market with a rescaling parameter γ_t for the inter-trade duration, and a rescaling parameter γ_x for the tick-by-tick logarithmic returns. Such parameters are either constants (γ_x), or obtained by draws of normally distributed random variables (γ_t, γ_x), and are reported in Table 3.

For each scenario, we present a realization of an intraday price series, the histogram of the final price as well as the payoff of an at-the-money intraday European call option, $C(T)$. This is given by $C(T) = \max(S(T) - K, 0)$, where T is maturity and K is the strike price; here, we assume $K = S(0)$. Figures 1–4 present Scenarios 1–4. Figure 5 shows the collective behavior of a portion of the market in the case of Scenario 2.

Table 1. The simulated scenarios: waiting times and jumps distributions.

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
Inter-trade durations J_t	Mittag-Leffler			ACD(1,1)
Log-returns Y_t	Lévy α-stable	ARCH(1)	GARCH(1,1)	

Table 2. Parameters for the ACD(1,1), ARCH(1) and GARCH(1,1) distributions.

	β	α_0	α_1	β_1
ACD(1,1)	1	$N(0.21, 0.01)$	$N(0.46, 0.01)$	$N(0.077, 0.001)$
ARCH(1)	0.99	$N(1, 0.1)$	$N(0.1, 0.01)$	-
GARCH(1,1)	1	$N(1, 0.1)$	$N(0.1, 0.01)$	$N(0.1, 0.01)$

3. Discussion and outlook

Using a simple set of equations and an appropriate set of stochastic processes, it is easy to simulate synthetic equity markets using Monte Carlo

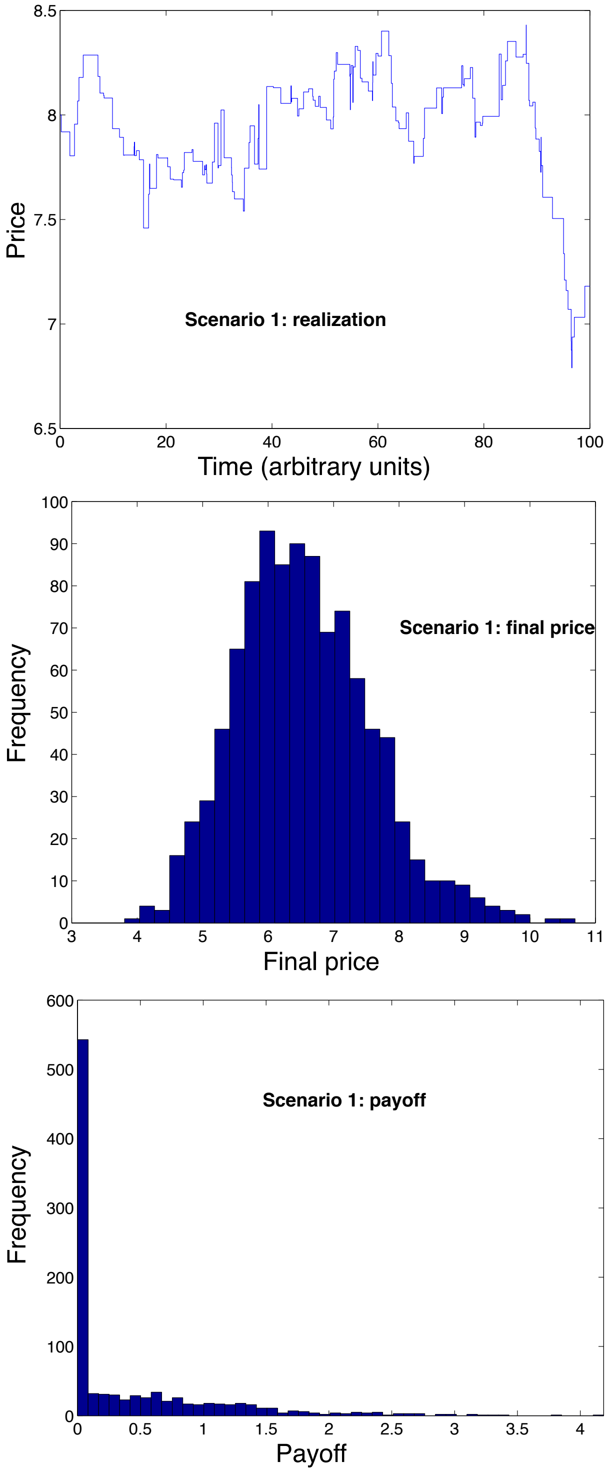

Figure 1. Top to bottom: a single share realization (time in arbitrary units), final price and payoff histograms for Scenario 1.

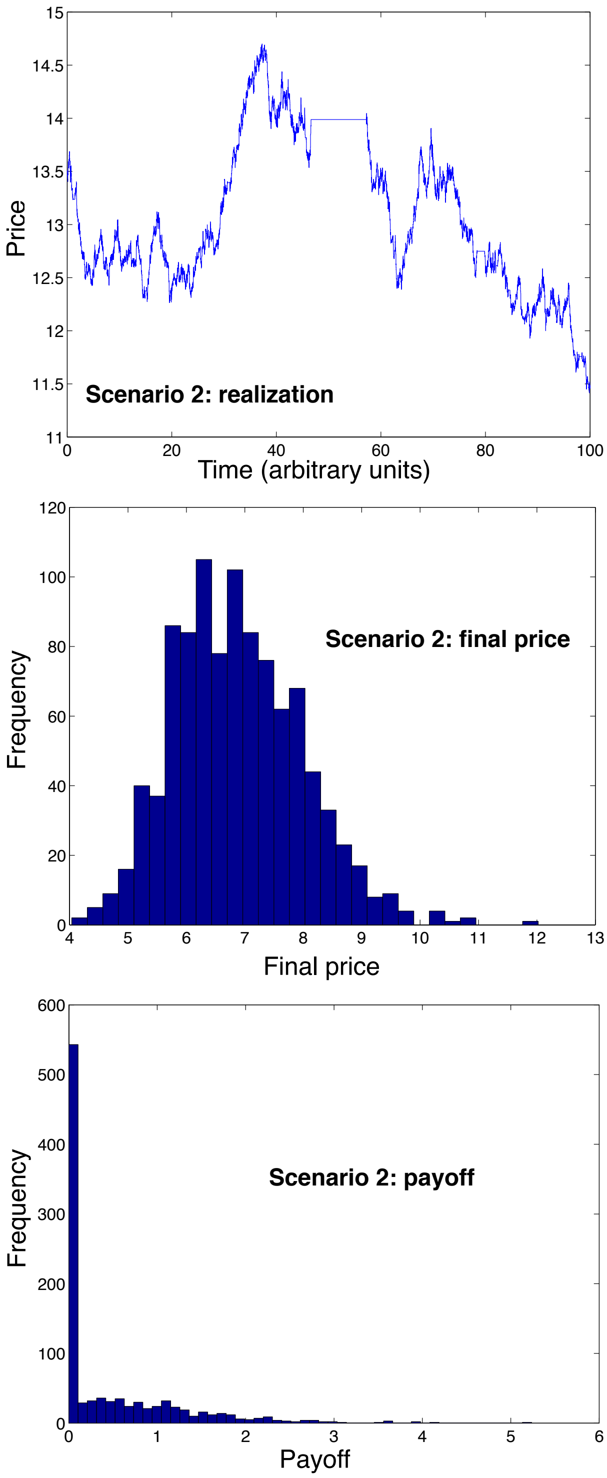

Figure 2. Top to bottom: a single share realization (time in arbitrary units), final price and payoff histograms for Scenario 2.

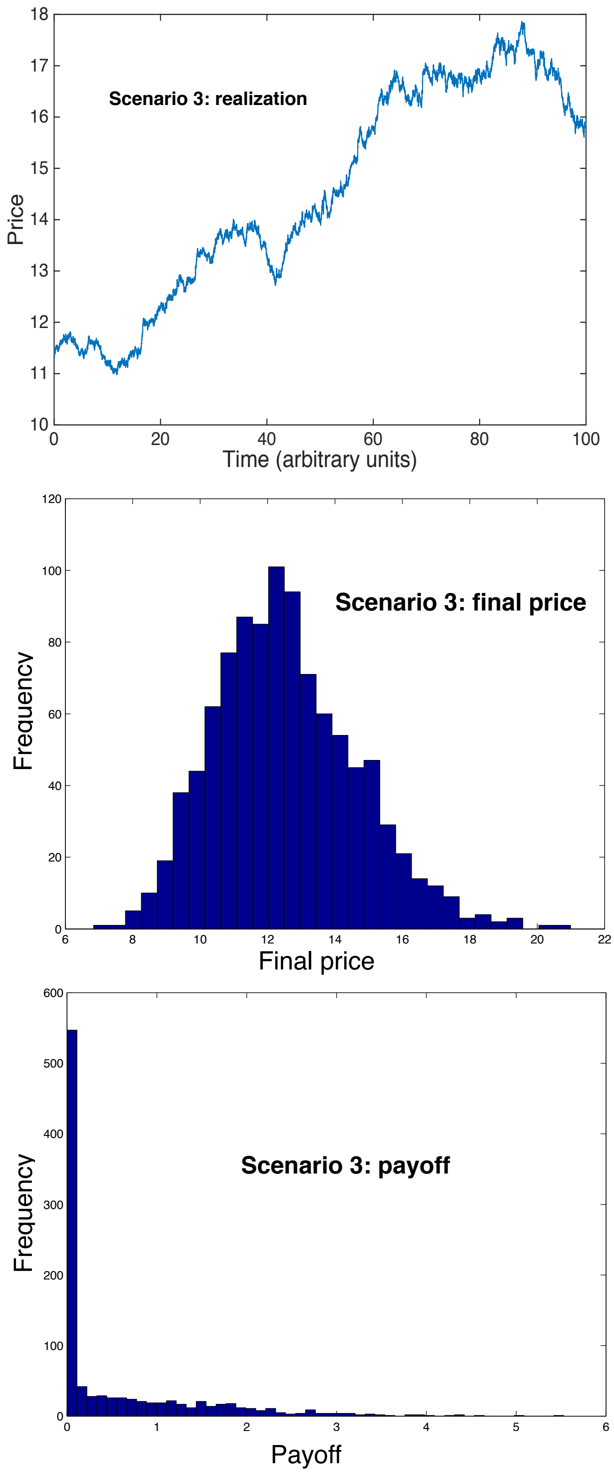

Figure 3. Top to bottom: a single share realization (time in arbitrary units), final price and payoff histograms for Scenario 3.

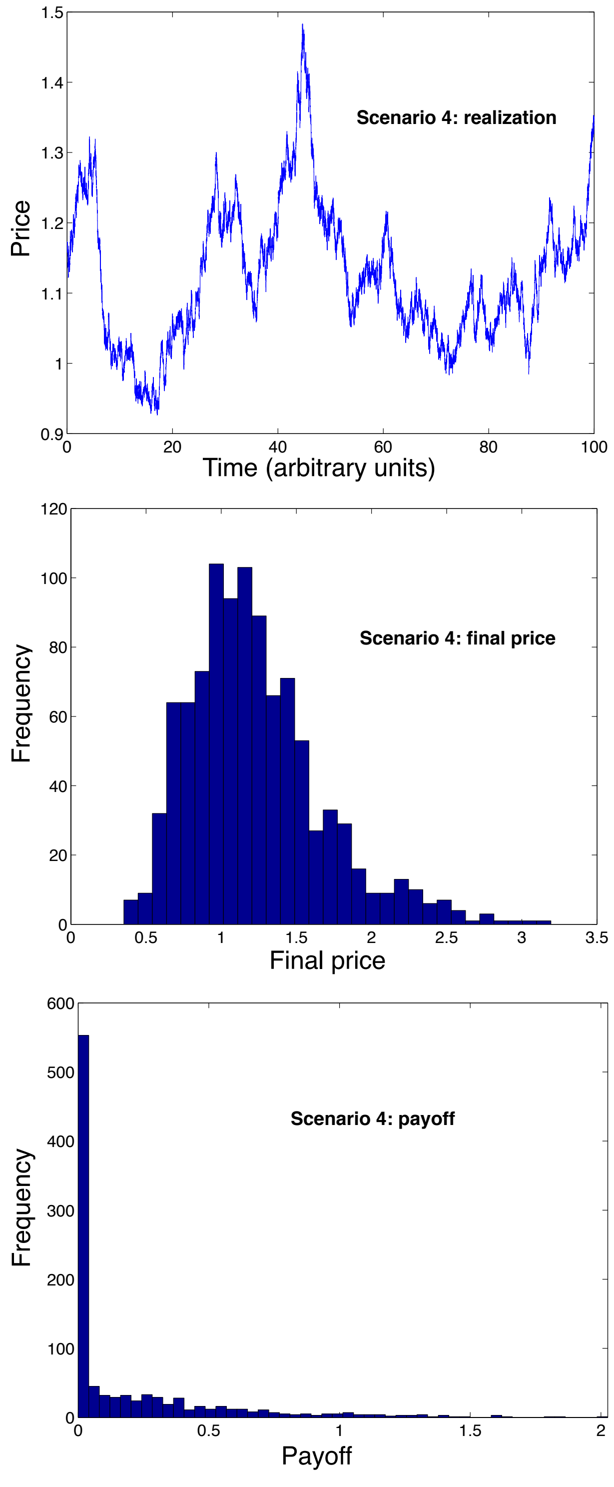

Figure 4. Top to bottom: a single share realization (time in arbitrary units), final price and payoff histograms for Scenario 4.

Table 3. The scale factors γ_t and γ_x.

	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**
γ_t	$N(0.01, 0.001)$	$N(0.01, 0.001)$	$N(0.01, 0.001)$	$N(0.01, 0.001)$
γ_x	$N(10^{-7}, 10^{-8})$	$10^{-5} \times \gamma_t$	10^{-7}	10^{-7}

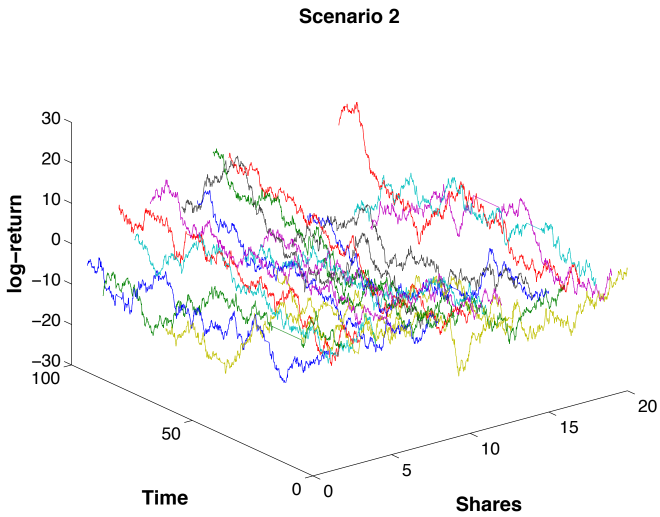

Figure 5. A sample showing the behaviour of 20 shares out of the market generated in Scenario 2

methods. Essentially one needs to generate a set of inter-trade durations and a set of tick-by-tick log returns. This can be easily done and the embarrassingly parallel nature of Monte Carlo simulations can be used to speed up the procedure if many processors are available.

We discussed a general 4-step procedure to build effective synthetic markets for high-frequency trading applications. The procedure includes model selection, fitting, simulations and assessment of quantities of interest. Here, we have only considered simulations and assessment of quantities of interest with some simple examples. We plan to devote future efforts on the more difficult problems of model selection and fitting. For fitting, we plan to use Bayesian hybrid Monte Carlo methods, whereas information criteria will be used for model selection.

Acknowledgements

L. Gerardo-Giorda was supported through the BERC 2014-2017 program of the Basque Government and the BCAM Severo Ochoa excellence accreditation from the Ministerio de Economia y Competitividad of the

DOI: 10.1685/journal.caim.535

Spanish Government (SEV-2013-0323). G. Germano gratefully acknowledges the support of the Economic and Social Research Council (ESRC) in funding the Systemic Risk Centre (grant number ES/K002309/1). E. Scalas acknowledges a SDF start-up research grant from the University of Sussex.

REFERENCES

1. Baleanu D., Diethelm K., Scalas E., Trujillo J. J., Fractional Calculus Models and Numerical Methods, World Scientific, Singapore, 2012.
2. Engle, R. F., and Russell, J. R., Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. Econometrica **66**, 1127–1162, 1998.
3. Furse, C., Bond, P., Cliff, D., Goodhart, C., Houstoun, K., Linton, O., Zigrand, J.-P., The Future of Computer Trading in Financial Markets (2012) Final Project Report The Government Office for Science, London. URL http://www.bis.gov.uk/foresight/our-work/projects/current-projects/computer-trading.
4. Fulger D., Scalas E., Germano G., Monte Carlo simulation of uncoupled continuous-time random walks yielding a stochastic solution of the space-time fractional diffusion equation, Phys. Rev. E, **77**, 021122, 2008.
5. Germano G., Engel M., Scalas E., City@home: Monte Carlo derivative pricing distributed on networked computers, in Proceedings of the 1st International Workshop on Grid Technology for Financial Modeling and Simulation, Palermo, Italy, 3-4 February 2006, edited by Stefano Cozzini, Stefano d'Addona, Rosario Mantegna, PoS(GRID2006)011.
6. Hautsch N., Econometrics of Financial High-Frequency Data, Springer, Berlin, 2012.
7. Leber C., Geib B., Litz H., High Frequency Trading Acceleration using FPGAs, International Conference on Field Programmable Logic and Applications (FPL), IEEE Conference Publications, 2011.
8. Livan G., Alfarano S., Scalas E., Fine structure of spectral properties for random correlation matrices: An application to financial markets, Phys. Rev. E **84**, 016113, 2011.
9. Livan G., Inoue J., Scalas E., On the non-stationarity of financial time series: impact on optimal portfolio selection, J. Stat. Mech., P07025, 2012.
10. Preis T., GPU-computing in econophysics and statistical physics, Eur. Phys. J. Special Topics **194**, 87–119, 2011.
11. Radivojević T., Anselmi J., Scalas E., Ergodic transition in a simple model of the continuous double auction, PLoS ONE **9**, e88095, 2014.
12. Scalas E., Gorenflo R., Mainardi F., Uncoupled continuous-time random walks: Solution and limiting behavior of the master equation, Phys. Rev.

E **69**, 011107, 2004.

13. Scalas E., Kim K., The art of fitting financial time series with Lévy stable distributions, Journal of the Korean Physical Society **50**, 105–111, 2007.
14. Scalas E., A Class of CTRWs: Compound Fractional Poisson Processes, in Klafter J., Lim S. C., Metzler R., Fractional Dynamics: Recent Advances, World Scientific, Singapore, 2011.
15. R. Weron, Lévy-stable distributions revisited: tail index ¿ 2 does not exclude the Lévy stable regime, Int. J. Mod. Phys. C **12**, 209-223, 2001.