Modeling non-stationarities in high-frequency financial time series

Article (Accepted Version)

Modeling non-stationarities in high-frequency financial time series

Linda Ponta, Mailan Trinh, Marco Raberto, Enrico Scalas,
Silvano Cincotti

PII: S0378-4371(19)30067-6
DOI: https://doi.org/10.1016/j.physa.2019.01.069
Reference: PHYSA 20500

To appear in: *Physica A*

Please cite this article as: L. Ponta, M. Trinh, M. Raberto et al., Modeling non-stationarities in high-frequency financial time series, *Physica A* (2019), https://doi.org/10.1016/j.physa.2019.01.069

Highlights

- Scaling properties for financial returns are still approximately satisfied.
- A simple stochastic process can approximate intra-day returns.
- Model selection is possible using information criteria.

Modeling non-stationarities in high-frequency financial time series

Linda Ponta[a,b], Mailan Trinh[c], Marco Raberto[b], Enrico Scalas[c,d,*], Silvano Cincotti[b]

[a]*LIUC – Cattaneo University, Corso G. Matteotti 22, 21053 Castellanza (VA), Italia*
[b]*DIME – CINEF, Università degli studi di Genova, Via Opera Pia 15, 16145 Genova, Italia*
[c]*Department of Mathematics, School of Mathematical and Physical Sciences, University of Sussex, Brighton, UK*
[d]*BCAM - Basque Center for Applied Mathematics, Bilbao, Basque Country - Spain*

Abstract

We study tick-by-tick financial returns for the FTSE MIB index of the Italian Stock Exchange (Borsa Italiana). We confirm previously detected non-stationarities. Scaling properties reported before for other high-frequency financial data are only approximately valid. As a consequence of our empirical analyses, we propose a simple model for non-stationary returns, based on a non-homogeneous normal compound Poisson process. It turns out that our model can approximately reproduce several stylized facts of high-frequency financial time series. Moreover, using Monte Carlo simulations, we analyze order selection for this class of models using three information criteria: Akaike's information criterion (AIC), the Bayesian information criterion (BIC) and the Hannan-Quinn information criterion (HQ). For comparison, we perform a similar Monte Carlo experiment for the ACD (autoregressive conditional duration) model. Our results show that the information criteria work best for small parameter numbers for the compound Poisson type models, whereas for the ACD model the model selection procedure does not work well in certain cases.

Keywords: stochastic processes, information criteria, high-frequency finance

*Corresponding author

Email addresses: linda.ponta@unige.it (Linda Ponta), m.trinh@sussex.ac.uk (Mailan Trinh), marco.raberto@unige.it (Marco Raberto), e.scalas@sussex.ac.uk (Enrico Scalas), silvano.cincotti@unige.it (Silvano Cincotti)

Introduction

The rise in the availability of high-frequency financial data has led to an increase in the number of studies focusing on the areas of classification and modeling of financial markets at the ultra-high frequency level. The development of models able to reflect the various phenomena observed in real data is an important step towards a full understanding of the fundamental stochastic processes driving the market. The statistical properties of high-frequency financial data and market micro-structural properties were studied by means of different tools, including phenomenological models of price dynamics and agent-based market simulations (see [1–31]).

Various studies on high-frequency econometrics appeared in the literature using the autoregressive conditional duration (ACD) models (see [32–35]). Alternative stochastic models were also proposed, e.g., diffusive models, ARCH-GARCH models, stochastic volatility models, models based on fractional processes, models based on subordinate processes (see [36–42]) as well as models based on self-exciting processes of Hawkes type [43–45]. An important variable is the order imbalance. Many existing studies analyze order imbalances around specific events or over short periods of time. For example, in [46] order imbalances are analyzed around the October 1987 crash. Reference [47] analyzes how order imbalances change the relation between stock volatility and volume using data for about six months. A large body of research examines the effect of the bid-ask spread and the order impact on the short run behavior of prices (see [48–61]). Trading activity was measured by the average number of trades in unit time intervals in [62] and [63]. However, aggregating trades into time intervals of the same length may have influences on the analysis. For instance, if intervals are too short with respect to the average waiting time between consecutive trades, then every interval will contain either no point or a small number of points. On the contrary, if intervals are too long, aggregation of too many points may lead to loss of information on the time structure of the process. Moreover, in both cases one distorts the kurtosis of the return process (see [33]).

For the reasons mentioned above, the waiting-time (duration) between two consecutive transactions is an important empirical variable (see [10, 21–25, 64–66]). In the market, during a trading day, the activity is not constant (see [32, 33]) leading to fractal-time behavior (see [67, 68]). Indeed, as a consequence of the double auction mechanism, waiting times between two subsequent trades are themselves random variables (see [64, 69, 70]). They may also be correlated to returns (see [71]) as well as to traded volumes.

In the Physics literature, in order to investigate tick-by-tick financial time series, the continuous-time random walk (CTRW) was used (see [4, 64, 72–75]). It turned out that interorder and intertrade waiting-times are not exponentially distributed. Therefore, the jump process of tick-by-tick prices is non-Markovian (see [4, 64]). Bianco and Grigolini applied a new method to verify whether the intertrade waiting time process is a genuine renewal process (see [76–78]). This was assumed by the CTRW hypothesis in [4]. They found that intertrade waiting-times do follow a renewal process. 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 fractions of a second to a few minutes, depending on the specific stock and on the market considered. In [71], 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, as mentioned above, the unconditional survival probability (the complementary cumulative distribution function) of waiting times is not exponentially distributed (see [64, 79]), but is well fitted by a Weibull function (see [12, 32, 33, 71, 80, 81]).

The non-stationary character of financial time series has also been the object of recent studies in the Physics literature [69, 82–85].

Here, inspired by [86], and building on the results presented in [69], we propose a model based on non-homogeneous Poisson processes. The paper is organized as follows. Section 1 describes the data set. Section 2 describes the statistical analysis of the single assets and of the FTSE MIB index, respectively as well as the scaling analysis; Section 3 contains the bivariate analysis whereas Section 4 is devoted to the compound Poisson model, its order selection and the numerical results. A comparison with order selection performance for ACD models is presented in the same section. Section 5 relates our methodology and results to the literature in Mathematics. Finally, Section 6 presents the conclusions of this work. A visual map of the structure of this paper is presented in Figure 1.

1. Description of the data set

The data set includes high-frequency trades registered at Italian Stock Exchange (BIt or Borsa Italiana), from the 03rd of February 2011 to the 09th of March 2011. The data of February 14th 2011 are not used because, on that day, there were technical problems at BIt. Moreover, we have removed the data of the 21st of February, as well. In fact, on that day, there was a crash in the Italian market related to the events in Lybia (on the 15th of February, a rebellion against the Lybian government begun). We consider

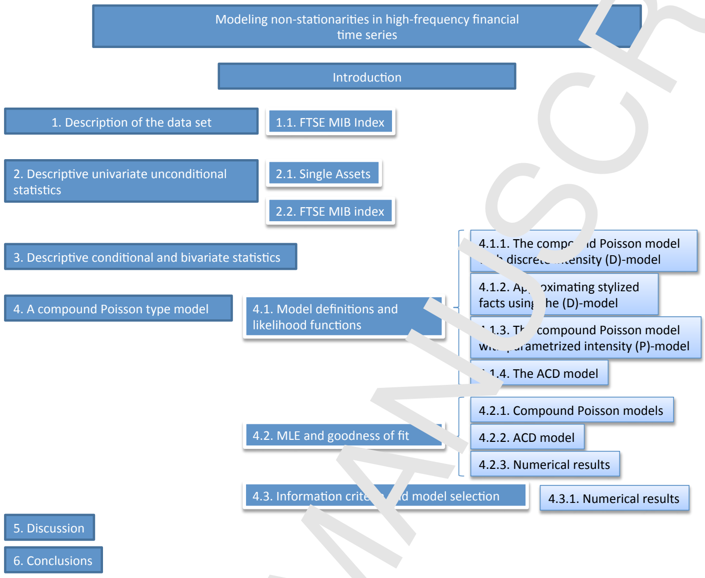

Figure 1: (Color online) Structure of the paper.

the 40 shares in the FTSE MIB Index as well as the index itself. Further information on the data set including the meaning of symbols and the calculation of the FTSE MIB index is available in the Supplemental material (see `https://github.com/enricoscalas/HFFnonstationary`). In particular, it is important to remark that the FTSE MIB Index value is updated every time there is a change of price of one of its components. The forty stocks composing the FTSE MIB vary in their average market capitalization and exhibit different levels of trading activity with different numbers of trades over this period as summarized in Table I in the Supplemental material where the total number of observations in the chosen month is given (see `https://github.com/enricoscalas/HFFnonstationary`). The number of data points per share varies between 10^4 and 10^5 and there are $4 \cdot 10^5$ values of the index. Choosing one month of high-frequency data was a trade-off between the necessity of using enough data for significant statistical analysis and, on the other hand, the goal of minimizing the effect of external economic fluctuations leading to non-stationarities of the kind discussed in [87]. For every stock, the data set consists of prices $p(t_i)$, volumes $v(t_i)$ and times of execution t_i sampled every second, where i is the trade index, varying from 1 to the total number of daily trades N. These data were filtered in order to remove misprints in prices and times of execution. In particular, concerning prices, when there are multiple prices for the same time of execution, we consider only one transaction at that time and a price equal to the average of the multiple prices. As far as waiting times, τ, between two executions are concerned, we remove observations larger than 200 s: This means more than 3 minutes without recorded trading.

1.1. FTSE MIB Index

The FTSE MIB Index (see [88]) is the primary benchmark index for the Italian equity markets. Capturing approximately 80% of the domestic market capitalisation, the Index is made up of highly liquid, leading companies across Industry Classification Benchmark (ICB) sectors in Italy. The FTSE MIB Index measures the performance of 40 shares listed on Borsa Italiana and seeks to replicate the broad sector weights of the Italian stock market. The Index is derived from the universe of stocks trading on BIt. The Index replaces the previous S&P/MIB Index, as a benchmark Index for Exchange Traded Funds (ETFs), and for tracking large capitalisation stocks in the Italian market. FTSE MIB Index is calculated on a real-time basis in EUR. The official opening and closing hours of the FTSE MIB Index series coincide with those of BIt markets and are 09:01 and 17:31 respectively. The FTSE MIB Index is calculated and published on all days when BIt is open for trading.

FTSE is responsible for the operation of the FTSE MIB Index. FTSE maintains records of the market capitalisation of all constituents and other shares and makes changes to the constituents and their weightings in accordance with the Ground Rules. FTSE carries out reviews and implement the resulting constituent changes as required by the Ground Rules. The FTSE MIB Index constituent shares are selected after analysis of the Italian equity universe, to ensure the Index best represents the Italian equity markets.

The FTSE MIB Index is calculated using a base-weighted aggregate methodology. This means the level of an Index reflects the total float-adjusted market value of all of the constituent stocks relative to a particular base period. The total market value of a company is determined by multiplying the price of its stock by the number of shares in issue (net of treasury shares) after float adjustment. An indexed number is used to represent the result of this calculation in order to make the value easier to work with and track over time. As mentioned above, the Index is computed in real time. The details on how to compute it can be found in [88].

2. Descriptive univariate unconditional statistics

In this section, we separately consider the descriptive univariate unconditional statistics for both the forty assets and for the FTSE MIB Index. By *univariate*, we mean that, here, we do not consider correlations between the variables under study. By *unconditional*, we mean that, here, we do not consider the non-stationary and seasonal behavior of the variables under study and the possible memory effects. Correlation and non-stationarity will be discussed in the next section.

2.1. Single Assets

In order to characterize market dynamics on a trade-by-trade level, we consider two variables: the series of time intervals between consecutive trades, τ and the trade-by-trade logarithmic returns, r. If $p(t_i)$ represents the price of a stock at time t_i where t_i is the epoch of the i-th trade, then we define the tick-by-tick log-return as

$$r_i = \log \frac{p(t_{i+1})}{p(t_i)}. \tag{1}$$

Note that $\tau_i = t_{i+1} - t_i$ is a random intertrade duration (and not a fixed time interval).

Among the empirical studies on τ, we mention [71, 89], concerning contemporary shares traded over a period of a few months, a study on rarely-traded nineteenth century shares in [90], and results on foreign exchange transactions in [91] and [92].

Tables 1 and 2 contain the descriptive statistics, evaluated for the entire sample, for the time series $\tau_i^h = t_{i+1}^h - t_i^h$ (with $t_0^h = 0$) and r_i^h, where the superscript h denotes the specific share and takes the label $h = I$ for the FTSE MIB Index.

In Table 1 the third and fourth columns give the two parameters of a Weibull distribution fit. The Weibull distribution has the following survival function:

$$\mathbb{P}(\tau > t) = P(t|\alpha, \beta) = \exp\left(-\alpha t^{\beta}\right), \tag{2}$$

where β is the shape parameter and α is the scale parameter. The values given in Table 1 were fitted using the moment method described in [70]. The quality of these fits is pictorially shown in Figure 2 for A2A, EXO, MS and TIT, respectively. The solid line represents our Weibull fit and the circles are the empirical data. Since different companies have different average intertrade duration $\langle \tau^h \rangle$ (see the second column in Table 1), they

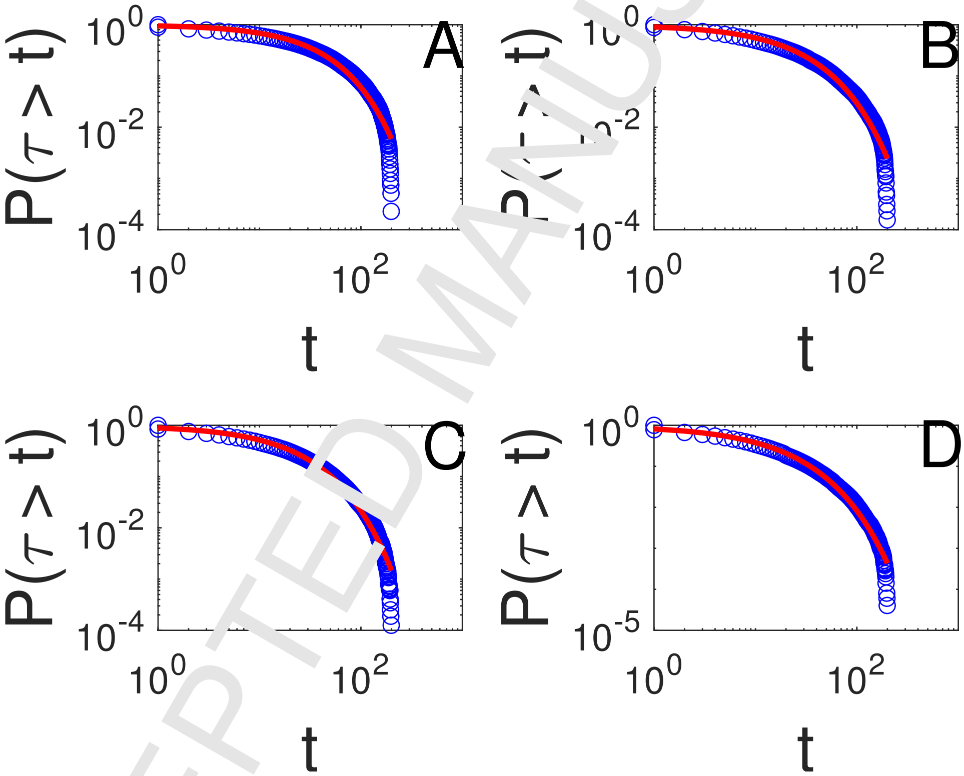

Figure 2: Weibull fit for A2A (A), EXO (B), MS (C), TIT (D). The fit is represented by the thin solid line, the open circles are the empirical values for the survival function.

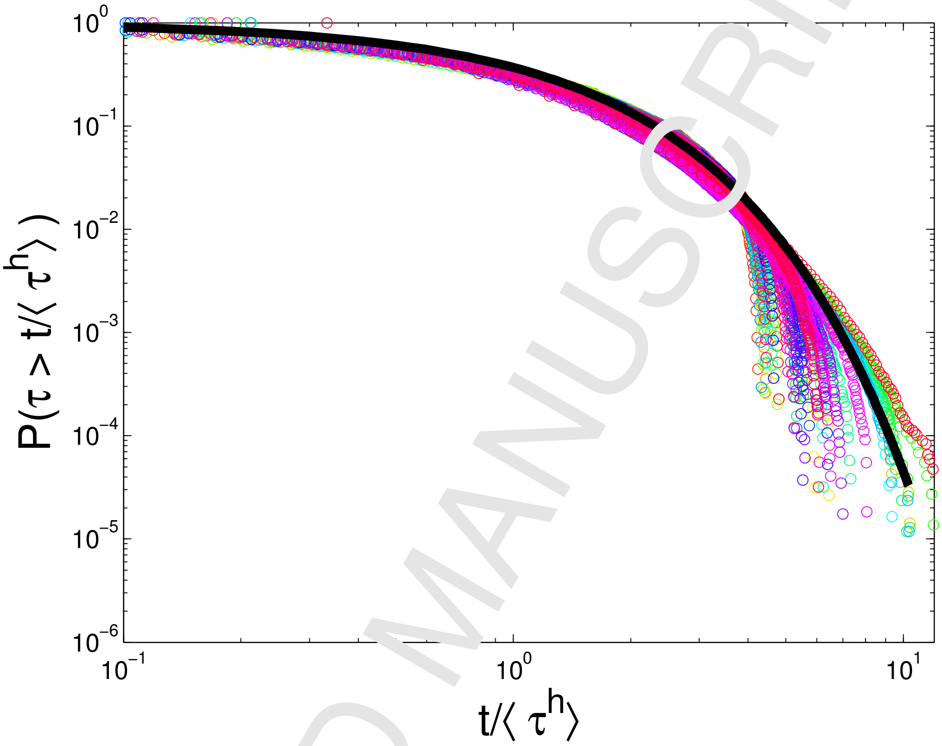

Figure 3: (Color online) Approximate scaling of the survival function for the forty time series. The solid line is the Weibull fit given by Eq.(3).

are also characterized by a different scale parameter α whereas the shape parameter β is almost the same for all the forty time series. Following [73], a scaling function $P(t|\beta^*)$ can be defined:

$$P(t|\beta^*) = \exp\left(-(t/\langle\tau\rangle)^{\beta^*}\right) \tag{3}$$

where $\beta^* = \langle\beta\rangle = 0.78$.

To test the hypothesis that there is a universal structure in the inter-trade time dynamics of different companies, we rescale the survival functions by plotting them against $t/\langle\tau^h\rangle$. We find that, for all companies, data approximately conform to a single scaled plot given by (3) as shown in Figure 3 (see also [70, 73, 93]). Such a behavior is a hallmark of scaling, and is typical of a wide class of physical systems with universal scaling properties [94]. Even if [95] showed that the scaling (3) is far from being universal, at least for the New York Stock Exchange, it is remarkable to find it again for a different index in a different market and seven

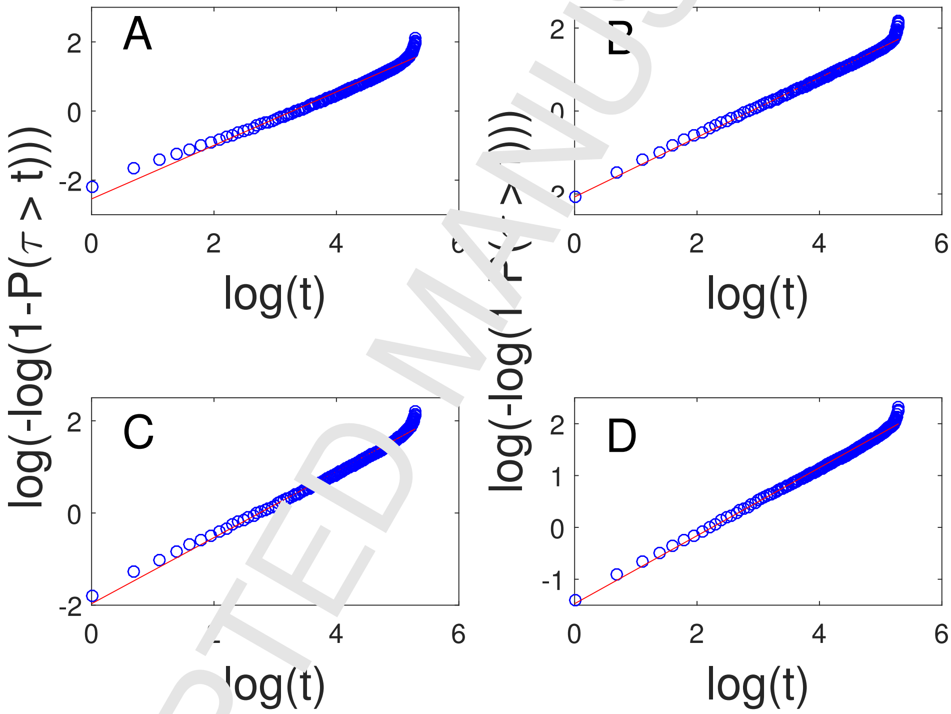

Figure 4: (Color online) Weibull paper for A2A (A), EXO (B), MS (C), TIT (D). On the horizontal axis, the values of $\log(t)$ are plotted, where t represents the inter-trade duration. On the vertical axes, a double logarithmic transform of the empirical cumulative distribution function of the inter-trade durations is plotted: $\log(-\log(1-P(\tau > t)))$. The linear fit is represented by the thin red solid line, the open circles are the empirical values.

years later with respect to the findings of [73]. However, to go beyond qualitative estimates, we perform several goodness-of-fit tests. Results for the Anderson-Darling and Lilliefors statistics are presented in Table 1. Results for the Kolmogorov-Smirnov test are in the Supplemental material (see `https://github.com/enricoscalas/HFFnonstationary`). All these tests reject the null hypothesis of Weibull distributed data. Finally, we present results based on the Weibull paper to graphically verify the Weibull distribution hypothesis. As an illustration, Figure 4 shows the Weibull paper for the following assets: A2A, EXO, MS and TIT. We can see that the deviation of the empirical data from the straight line expected for the Weibull distribution is mainly due to the tails of the distribution as expected from visual inspection of Figure 3.

The descriptive statistics for trade-by-trade returns r^h can be found in Table 2. Notice that there is excess kurtosis.

2.2. FTSE MIB index

We now investigate the FTSE MIB index. Tables 1 and 2 summarize also the descriptive statistics of the time series τ_i^I and r_i^I respectively evaluated for the FTSE MIB index.

In Figure 5 we show the survival function for the intertrade waiting time of the FTSE MIB index. The solid line represents the Weibull fit, whereas the circle represents the empirical data. The shape of the two curves is very different. Therefore, we can immediately see that intertrade times are not Weibull distributed, and, in this case, the fit does not work even as a first approximation. Indeed, for the FTSE MIB index, the standard deviation of intertrade durations is smaller than the average intertrade duration and the AD test and the Lilliefors test reject the null hypothesis of Weibull distribution.

Contrary to the case of single asset returns, the excess kurtosis for the FTSE MIB index is quite large. Figure 6 shows the histogram of the returns for a bin size of 1×10^{-5}.

Following [18], we test the scaling of the empirical returns. The dataset consists of 405560 records for the FTSE MIB index (Table I in the Supplemental Material `https://github.com/enricoscalas/HFFnonstationary`) during the period studied (from the 03rd of February 2011 to the 09th of March 2011). From this database, we compute the new random variable $r^I(t;\Delta t)$ defined as:

$$r^I(t;\Delta t) = \log \frac{p^I(t+\Delta t)}{p^I(t)}, \tag{4}$$

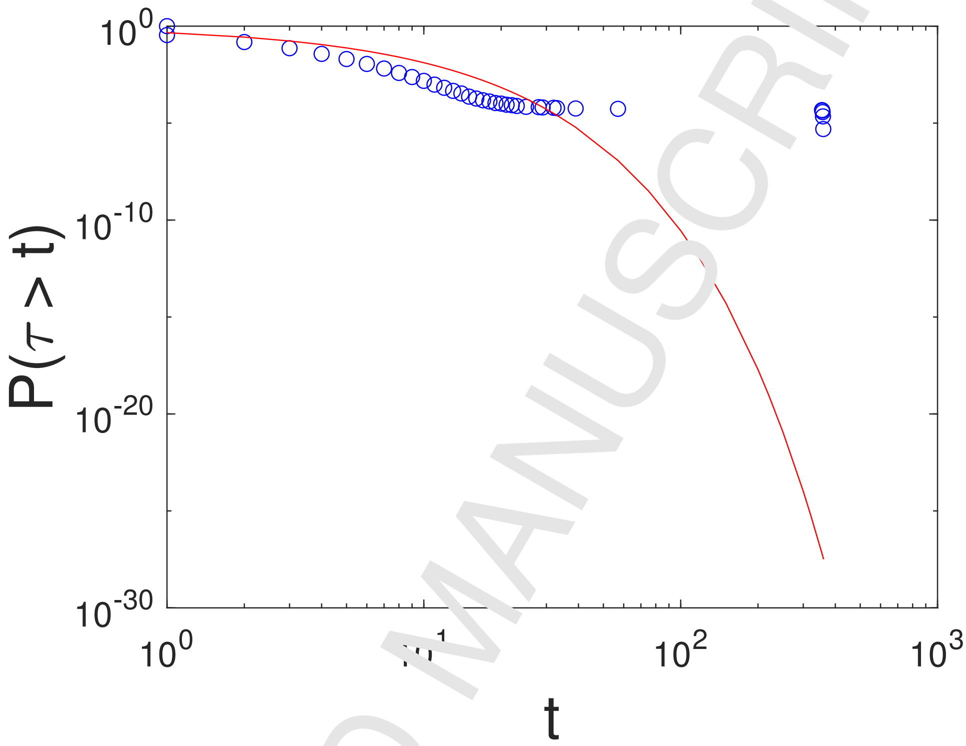

Figure 5: (Color online) Circles: empirical survival function; solid line: Weibull fit.

where $p^I(t)$ is the value of the index at time t. In this way we sample returns on equally spaced and non-overlapping intervals of width Δt. We further assume that the time series is stationary so that it only depends on Δt and not on t (incidentally, we shall later see that this is not the case). To characterize the experimentally observed process quantitatively, we first determine the empirical probability density function $P(r^I(\Delta t))$ of index variations for different values of Δt. We select Δt equal to 3s, 5s, 10s, 30s and 300s. In Figure 7(a) we present a semi-logarithmic plot of $P(r^I(\Delta t))$ for the five different values of Δt indicated above. These empirical distributions are roughly symmetric and are expected to converge to the normal distribution when Δt increases. The null hypothesis of normal distribution has been tested with the Kolmogorov-Smirnov, the Jarque-Bera and the Lilliefors test and is always rejected.

As already mentioned, we also note that the distributions are leptokurtic, that is, they have tails heavier than expected for a normal distribution. A determination of the parameters characterizing the distributions is difficult

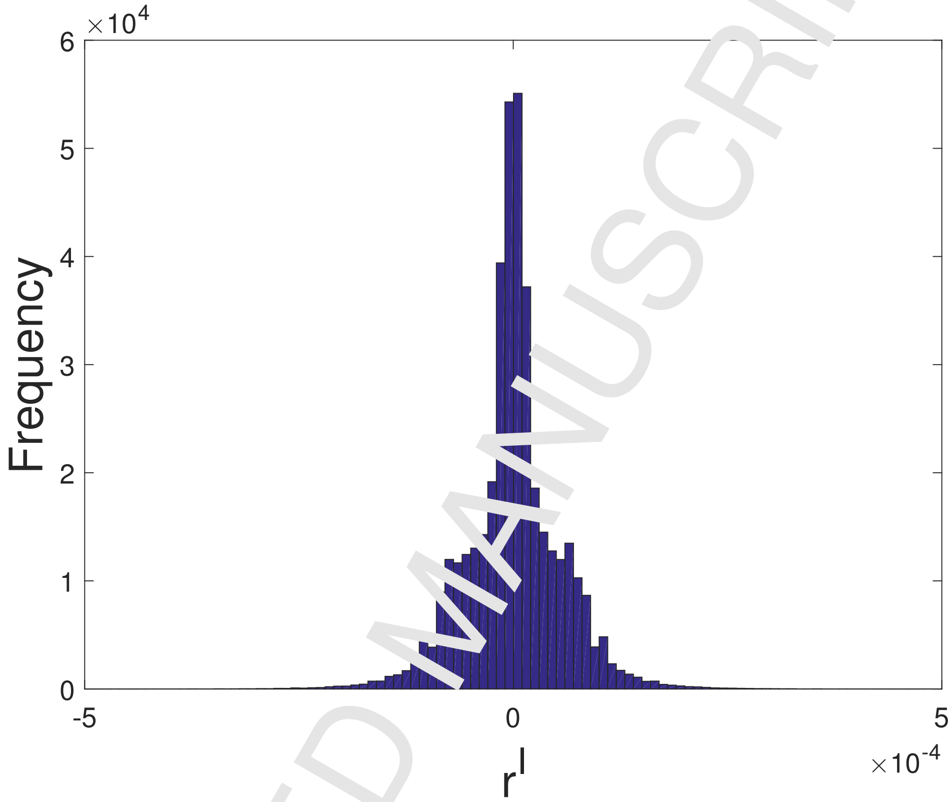

Figure 6: (Color online) Histogram of returns for the FTSE MIB index.

especially because larger values of Δt imply a smaller number of data. Again following [18], we study the probability density at zero return $P(r^I(\Delta t) = 0)$ as function of Δt. This is done in Figure 7(b), where $P(r^I(\Delta t) = 0)$ versus Δt is shown in a log-log plot. If these data were distributed according to a symmetric α-stable distribution, one would expect the following form for $P(r^I(\Delta t) = 0)$ (see Equation (2) in [18]):

$$P(r^I(\Delta t) = 0) = \frac{\Gamma(1/\alpha_L)}{\pi \alpha_L (c\Delta t)^{1/\alpha_L}}, \tag{5}$$

where $\Gamma(\cdot)$ is Euler Gamma function, $\alpha_L \in (0, 2]$ is the index of the symmetric α-stable distribution and c is a time-scale parameter. The data are well fitted (in the OLS sense) by a straight line of slope $1/\hat{\alpha}_L = 0.58$ leading to an estimated exponent $\hat{\alpha}_L = 1.72$. The best method to get the values of $P(r^I(\Delta t) = 0)$ is to determine the slope of the cumulative distribution

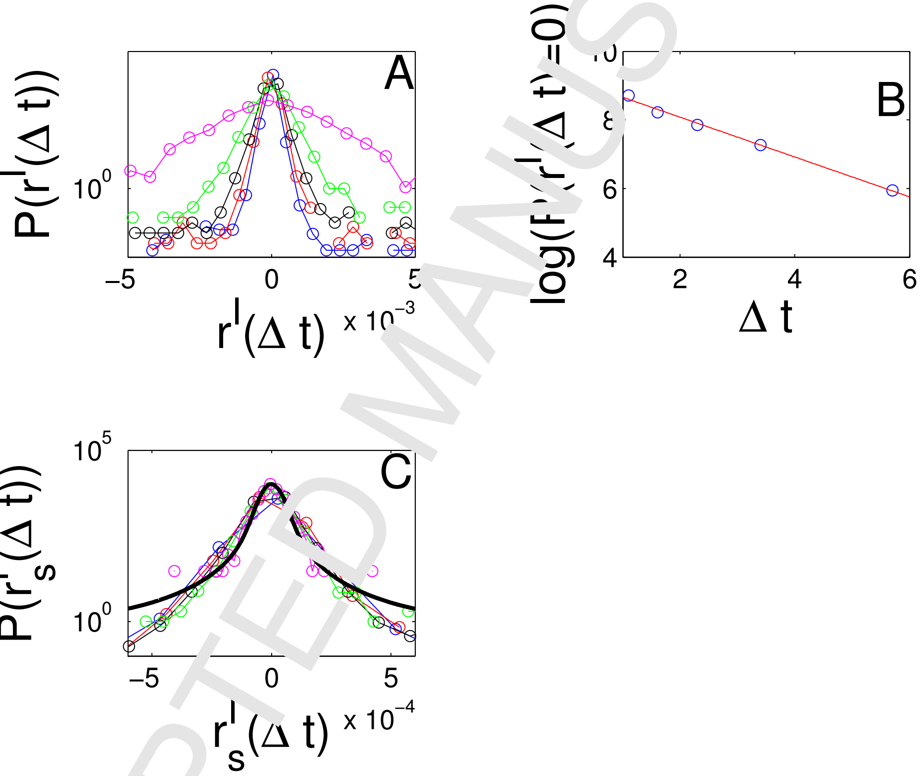

Figure 7: (Color online) (A) Histogram of the returns for the FTSE MIB index observed at different time intervals, namely, Δt = 3 s (blue), 5 s (red), 10 s (black), 30 s (green) and 300 s (purple); (B) Probability of zero returns as a function of the time sampling interval Δt, the slope of the straight line is 0.58 ± 0.01; (C) scaled empirical probability distribution and comparison with the theoretical prediction given by Eq.(7) (black solid line).

function in $r^I(\Delta t) = 0$. In Figure 7(c), we plot the rescaled probability density function according to the following transformation:

$$r_s^I = \frac{r^I(\Delta t)}{(\Delta t)^{1/\alpha_L}} \tag{6}$$

and

$$P(r_s^I) = \frac{P(r^I(\Delta t))}{(\Delta t)^{-1/\alpha_L}}, \tag{7}$$

for $\alpha_L = \hat{\alpha}_L = 1.72$. Remarkably all the five distributions approximately collapse into a single one. We use the Kolmogorov-Smirnov test to study the null hypothesis of identically distributed rescaled data; the results are shown in Table 3. The null hypothesis is rejected only in the following cases: $\Delta t = 3s$ and $\Delta t = 5s$, $\Delta t = 3s$ and $\Delta t = 10s$, $\Delta t = 3s$ and $\Delta t = 30s$.

It is worth noting that this result shows that the scaling, found in the S&P 500 data by Mantegna and Stanley more than twenty years ago [18], still approximately holds in a different market and in a completely different period. We do not run hypothesis tests on the Lévy stable distribution because an eye inspection of Figure 7(c) is sufficient to conclude that the Lévy stable fit is not matching the rescaled data.

3. Descriptive conditional and bivariate statistics

Inspired by [86, 96], in order to study the time variations of the returns during a typical trading day, we use a simple technique. We divide the trading day into equally spaced and non-overlapping intervals of length δt for $\delta t = 3, 5, 10, 30, 300, 600, 900, 1200, 1500$ and 1800 s. Let K be number of intervals and N_k the number of transaction in each interval k. For each interval we evaluate the $\gamma(k)$ indicator as a measure of volatility. $\gamma(k)$ is defined as

$$\gamma(k) = \frac{1}{N_k - 1} \sum_{i=1}^{N_k - 1} |r_{k,i}^I - \langle r_k^I \rangle|; \tag{8}$$

where $\langle r_k^I \rangle$ is the average value of returns in the time interval k. In Figure 8(a), as an example, we plot the average value of $\gamma(k)$ over the investigated period as a function of the interval index k for $\delta t = 300$ s. We can see that the volatility is higher in the morning, at the opening of continuous trading, and then it decreases up to midday. There is a local increase after midday and then the volatility returns to lower values to finally grow towards the end of continuous trading. In Figure 8(b), we plot the number of trades on

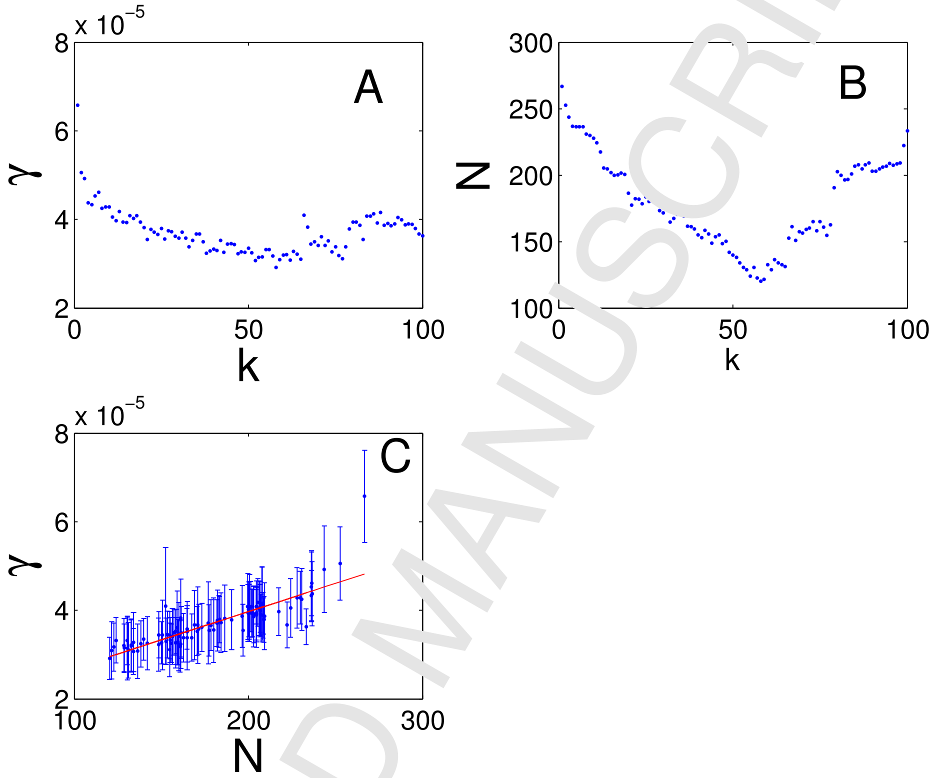

Figure 8: (Color online) (A) Volatility γ as a function of k for $\delta t = 300$ s. (B) Activity N as a function of k for $\delta t = 300$ s. (C) Scatter plot of volatility γ as a function of number of trades N. The points are averaged over the investigated period.

the FTSE MIB index as a function of the interval index k for $\delta t = 300$ s. The behavior of the trade activity closely follows the behavior of volatility. This is even clearer from the analysis of Figure 8(c) where the volatility is plotted as a function of the activity. The scatter plot shows a strong correlation between the two variables. This result does not depend on the length of the interval w, but the corresponding plots are not presented here for the sake of compactness. This feature was already present in the Australian market studied for a much longer period (10 years $\approx$ 2500 days) by [86, 96]. Again, it is remarkable to see a statistical pattern still valid in a different market after more than 10 years.

Figure 8 shows a seasonal pattern in intraday trades. In order to take this behavior into account, we proposed to use a non-stationary normal compound Poisson process with volatility of jumps proportional to the activity of the Poisson process in [69]. Here, we take even a more pragmatic stand and we do not assume any *a priori* relationship between volatility and activity as it emerges spontaneously, if present, with the method described in the next section.

4. A compound Poisson type model

As one can see, during a trading day, the volatility and the activity are higher at the opening of the market, then they decrease at midday and they increase again towards market closure [96] (see also Figure 8). In other words, the (log-)price process is non-stationary. As suggested in [69], such a non-stationary process for log-prices can be approximated by a mixture of normal compound Poisson processes (NCPP) in the following way. A normal compound Poisson process is a compound Poisson process with normal jumps. In formula:

$$X(t) = \sum_{i=1}^{N(t)} R_i, \tag{9}$$

where R_i are normally distributed independent trade-by-trade log-returns, $N(t)$ is a Poisson process with parameter λ and $X(t)$ is the logarithmic price, $X(t) = \log(P(t))$. By probabilistic arguments one can derive the cumulative distribution function of $X(t)$, it is given by:

$$F_{X(t)}(u) = \mathbb{P}(X(t) \leq u) = \mathrm{e}^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} F_R^{\star n}(u), \tag{10}$$

where $F_R^{\star n}(u)$ is the n-fold convolution of the normal distribution, namely

$$F_R^{\star n}(u) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{u - n\mu}{\sqrt{2n\sigma^2}}\right)\right], \tag{11}$$

and μ and σ^2 are the parameters of the normal distribution.

We now assume that the trading day can be divided into n equal intervals of constant activity $\{\lambda_i\}_{i=1}^n$ and of length w, then the unconditional waiting time distribution becomes a mixture of exponential distributions and its cumulative distribution function can be written as

$$F_\tau(u) = \mathbb{P}(\tau \leq u) = \sum_{i=1}^{n} a_i (1 - \mathrm{e}^{-\lambda_i \tau}), \tag{12}$$

where $\{a_i\}_{i=1}^n$ is a set of suitable weights. The activity seasonality can be mimicked by values of λ_i that decrease towards midday and then increase again towards market closure. In order to reproduce the correlation between volatility and activity, one could assume that

$$\sigma_{\xi,i} = c\lambda_i \tag{13}$$

where c is a suitable constant. As already mentioned, however, for practical purposes, one can also estimate three parameters for each interval, the parameter λ_i of the Poisson process and the parameters μ_i and σ_i for the log-returns without any correlation assumption. This leads us to two possible examples of such compound Poisson type models which will be introduced in Section 4.1 alongside the popular ACD model for later comparisons. After a brief error analysis of the maximum likelihood estimation (MLE) method in Section 4.2, we will move on to the main Monte Carlo experiment to test model selection using information criteria (IC) in Section 4.3. The different nature of the compound Poisson models and the ACD model makes a direct comparison in terms of model selection questionable. Therefore, our main focus will be a comparison of IC within each model class separately.

4.1. Model definitions and likelihood functions

4.1.1. The compound Poisson model with discrete intensity (Dλ)-model

We extend the notation of Equation (9) by an additional index denoting the corresponding interval: We suppose that high-frequency data is given over a time interval $[t_0, T]$. First, set a time grid $\{t_i\}_{i\in\{1,\ldots,n\}}$ such that $t_0 < t_1 < t_2 < \ldots < t_n = T$. On each time interval $[t_{i-1}, t_i)$ we have a compound Poisson process

$$X_i(t) := \sum_{k=1}^{N_i(t)} R_k^{(i)}, \tag{14}$$

where $\{R_k^{(i)}\}_{k\in\mathbb{N}}$ is an i.i.d. sequence of $\mathcal{N}(\mu_i, \sigma_i^2)$ distributed random variables and $(N_i(t))_{t\geq 0}$ is a homogeneous Poisson process with parameter λ_i. Further, $\{R_k^{(i)}\}_{k\in\mathbb{N}}$ are all independent of $(N_i(t))_{t\geq 0}$.
For a fixed time interval $[t_{i-1}, t_i)$ the log-likelihood function is given by

$$\begin{aligned}\mathcal{L}_i^D(\lambda_i, \mu_i, \sigma_i) =& -\lambda_i(t_i - t_{i-1}) + \ln(\lambda_i)N_i(t_i)\\ &+ \sum_{k=1}^{N_i(t_i)} \ln(p_{\mu_i,\sigma_i}(R_k^{(i)})),\end{aligned} \tag{15}$$

where p_{μ_i,σ_i} denotes the probability density function of the $\mathcal{N}(\mu_i, \sigma_i^2)$ distribution. Due to the independence assumptions the overall log-likelihood is given by the sum of all $\mathcal{L}_i$. Equation (15) can be derived from the general expression for the sample density function given on page 200 in [97] by substituting a constant λ.
The maximum likelihood estimators are therefore:

$$\hat{\lambda}_i = N_i/w_i, \quad \hat{\mu}_i = \frac{1}{N_i}\sum_{k=1}^{N_i} r_i,$$

$$\hat{\sigma}_i^2 = \frac{1}{N_i}\sum_{k=1}^{N_i}(r_i - \hat{\mu}_i)^2, \tag{16}$$

where N_i is the number of trades in the ith interval and $w_i = t_i - t_{i-1}$.
Note that the maximum likelihood estimator for σ^2 is biased and the bias can be corrected by using

$$\tilde{\sigma}_i^2 = \frac{1}{N_i - 1}\sum_{k=1}^{N_i}(r_i - \hat{\mu}_i)^2 \tag{17}$$

instead. We shall use either the biased or unbiased estimator in the following sections when appropriate.

4.1.2. Approximating stylized facts using the (Dλ)-model

A Monte Carlo simulation of the (Dλ)-model was performed by considering a trading day divided into a number of intervals of length $w = \Delta t = 3, 5, 10, 30, 300$ s. The parameters $\hat{\lambda}_i$, $\hat{\mu}_i$ and $\tilde{\sigma}_i^2$ were estimated as explained above. Note that we use the unbiased estimator $\tilde{\sigma}_i$ from (17). In the following, we shall focus on estimates based on the FTSE MIB index. In Figure 9, we empirically show that the simulation gives a better fit for the empirical returns of the index as w becomes smaller. This figure corroborates the conjecture that the approximations converge to the empirical data. This is an encouraging result meaning that it will be useful to study the convergence of the approximation by means of measure-theoretical probabilistic methods. Figure 10 displays the histogram of simulated returns for $w = 3$ and can be compared to Figure 6. The corresponding value of the Kolmogorov-Smirnov statistics is given by the blue dot in Figure 9.

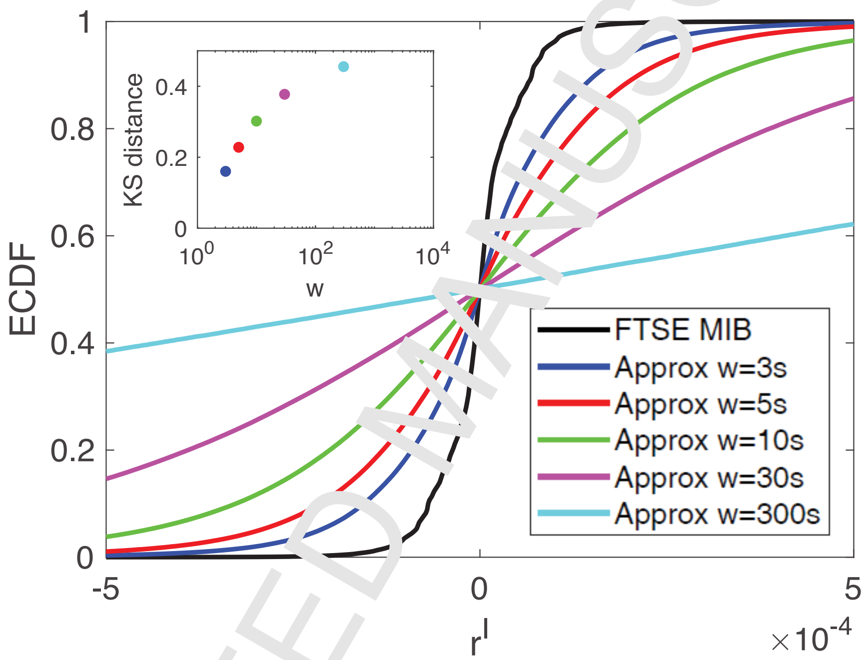

Figure 9: (Color online) Approximation of the empirical cumulative distribution function with Monte Carlo simulations for FTSE MIB returns r^I. The black line represents the empirical cumulative distribution functions for real data. The colored lines represent the simulations described in the text and based on sampling at equal intervals of 3, 5, 10, 30 and 300 seconds as described by the legend. The inset contains a plot of the Kolmogorov-Smirnov distance between the approximations and the empirical curve. This plot corroborates the conjecture that there is convergence of the approximation to the black curve.

In order to show that this approximation is able to reproduce the approximate stylized facts described above, Figure 11 shows the scaling relations discussed in section 2.2 for the simulation with $w = 10$ s. The null hypothesis of normal distribution has been tested with the Kolmogorov-Smirnov, the Jarque-Bera and the Lilliefors test. Also in this case the null hypothesis is always rejected.

One can see from Figure 11(b) that an OLS index estimate $\hat{\alpha}_L = 1.59$ is

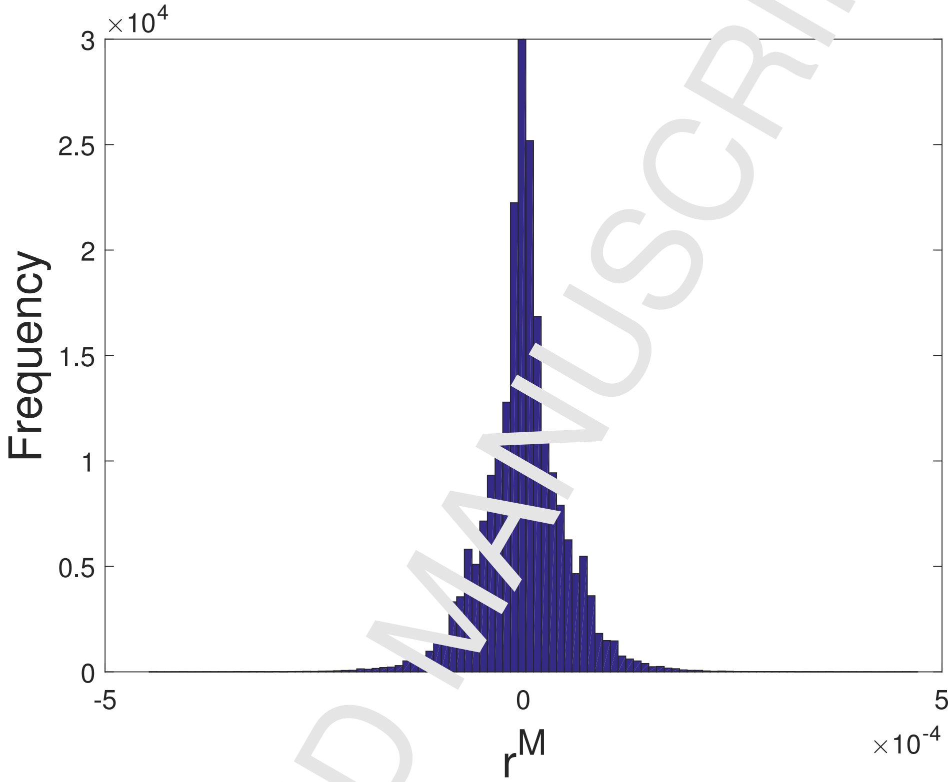

Figure 10: (Color online) Histogram of returns for the approximating process with $w = 3$s.

recovered from the simulation instead of 1.72 for the real index. The scaling given in Eqs. (6), (7) is presented in Figure 11(c), one can see that the approximate scaling still holds for the simulated data. The null hypothesis of identical distribution has been tested with the Kolmogorov-Smirnov test, and the results are shown in Table 4. It is worth noting that the null hypothesis of identical distribution is always rejected but the statistic value is near to the critical value.

4.1.3. The compound Poisson model with parametrized intensity (Pλ)-model

This model will be used for simulation later on as well as serve as a benchmark model when testing model selection criteria. As empirical results about the trading intensity suggest a daily seasonality, this model assumes that the step function in the (Dλ) model is parametrized by a quadratic

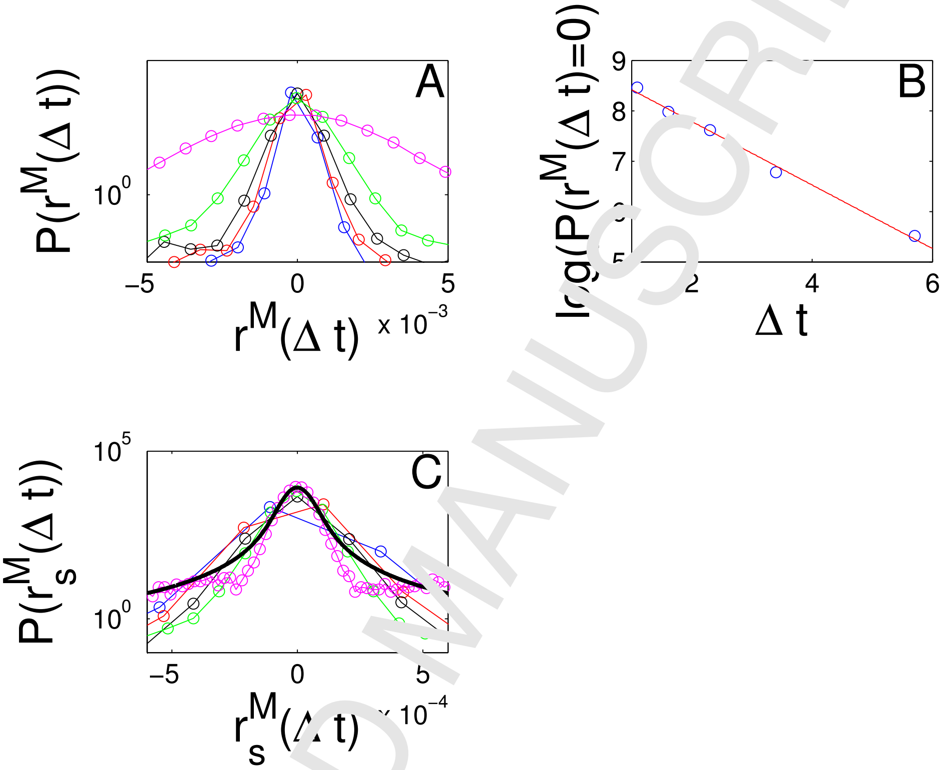

Figure 11: (Color online) (A) Histogram of the returns for the simulation described in the text observed at different time intervals, namely, $\Delta t = 3$ s (blue), 5 s (red), 10 s (black), 30 s (green) and 300 s (purple); (B) Probability of zero returns as a function of the time sampling interval Δt, the slope of the straight line is 0.63 ± 0.01; (C) scaled empirical probability distribution and comparison with the theoretical prediction given by Eq.(7) (black solid line).

function:

$$\lambda_{a,b,c}(t) = at^2 + bt + c, \quad t \in [0, 1]. \tag{18}$$

Of course, this parametrization can be easily replaced by more complicated functions. Since λ needs to be positive and convex, we also have the conditions

$$a > 0 \text{ and } c > \frac{b^2}{4a}. \tag{19}$$

Similar to the (Dλ)-model, the log-likelihood for the (Pλ)-model is given by

$$\mathcal{L}_i^P(a,b,c,\mu_i,\sigma_i) = -\lambda_{a,b,c}(t_{i-1})(t_i - t_{i-1}) + \ln(\lambda_{a,b,c}(t_{i-1}))N_i(t_i) + \sum_{k=1}^{N_i(t_i)} \ln(p_{\mu_i,\sigma_i}(R_k^{(i)})). \tag{20}$$

While the maximum likelihood estimators for μ_i and σ_i are the same as for the (Dλ) case, the maximum likelihood estimators for a, b, c, which determine the form of λ, cannot be obtained in closed form. As a consequence, a numerical optimization method needs to be applied to estimate those parameters.

4.1.4. The ACD model

The autoregressive conditional duration model was first proposed by Engle and Russell [33]. We will consider a model for the durations between events only without marks: Let $(\varepsilon_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables. The autoregressive conditional duration (ACD) model is defined as

$$x_i = \psi_i \varepsilon_i \tag{21}$$
$$\psi_i \equiv \psi_i(x_{i-1},\ldots,x_1;\theta) := \mathbb{E}\left[x_i | x_{i-1},\ldots,x_1\right]. \tag{22}$$

The innovations (ε_i) are assumed to follow an exponential distribution, i.e. $\varepsilon_i \sim \text{Exp}(1)$, and ψ_i has the following representation

$$\psi_i := \omega + \sum_{j=0}^{m} \alpha_j x_{i-j} + \sum_{j=0}^{q} \beta_j \psi_{i-j}, \tag{23}$$

where $\omega > 0$, $\alpha_i \geq 0$ and $\beta_i \geq 0$ for all i. We will call this model ACD(m,q). For given duration data $\{x_1,\ldots,x_n\}$ the log-likelihood function is given by

$$\mathcal{L}^{\text{ACD}}(\omega,\alpha_1,\ldots,\alpha_m,\beta_1,\ldots,\beta_q) = -\sum_{i=1}^{n}\left[\ln\psi_i + \frac{x_i}{\psi_i}\right] \tag{24}$$

(see p. 104 in [20]).

4.2. MLE and goodness of fit

Before we turn our attention to the actual model selection procedure, it is useful to get a rough idea about how well the underlying MLE method works for the three model classes. We would like to ensure that the MLE method works reasonably well since a poor ML fit might compromise the quality of the order selection. Due to asymptotic results, we expect that goodness of fit and correctness of the model selection procedure should improve with increasing size of the underlying sample. As these two effects are closely related, it is hard to quantify them separately.
In the next sections, we give a detailed explanation of the simulation procedure and on how the parameter estimation is implemented. Based on that, we run a MLE on previously generated mock data. As we know the true parameter values, we can easily calculate the mean squared error (MSE) as measure for the goodness of fit.

4.2.1. Compound Poisson model
Simulation. The simulation algorithm essentially uses the (Pλ)-model. For simplicity we will choose the time interval $[t_0, T]$ to be $[0, 1]$. For the simulation we set an equidistant grid $0 = t_0 < t_1 < t_2 < \ldots < t_n = 1$ on the time interval. Thus, the interval $[0, 1]$ is divided into n subintervals. For $i \in \{1, \ldots, n\}$ the parameters μ_i, σ_i and λ_i on the subinterval $[t_{i-1}, t_i)$ are chosen to be

$$\begin{aligned} &\mu_i = 0,\ \sigma_i = 1 \text{ and } \lambda_i = \lambda(t_{i-1})\ \forall i \in \{1, \ldots, n\}, \\ &\text{where } \lambda(t) := 4(\lambda_{\max} - \lambda_{\min})(t - 0.5)^2 + \lambda_{\min}, \\ &\quad \forall t \in [0, 1] \text{ and } \lambda_{\min}, \lambda_{\max} > 0 \text{ constant.} \end{aligned} \tag{25}$$

The functional form of λ is inspired by the empirical findings in the previous sections and should account for the observed seasonality in a simple way. We have chosen $\lambda_{\min} = 100$ and $\lambda_{\max} = 10000$. Note that the $\{\lambda_i\}$ form a step function approximation of the parabola in (25). For different grid sizes, we simulate with sample size 1000 each.

Fitting. The fitting is carried out using different grid sizes. Note that the grid size to be used in fitting is bounded from above by the length of the entire time interval (in our case 1). However, we would like to emulate the behavior of the intensity which was observed in empirical data, i.e. high intensity at the beginning and at the end of the trading day and relatively low intensity in the middle of the day. Consequently, we need at least 3 subintervals to have a piecewise constant function that fulfils these conditions on

the time interval. Further, the smallest eligible grid size is bounded from below by the maximal distance between neighbouring data points within the data set. Otherwise, there are subintervals which do not contain any data points. In such cases, the estimation formulas in (16) would fail.
More precisely, for the maximal distance $\Delta_{\max}$ between two consecutive data points within a given sample, the finest valid equidistant grid has at most $\left\lfloor \frac{1}{\Delta_{\max}} \right\rfloor$ subintervals. Therefore, we will consider a list of candidate models on grids which correspond to $n = 3, 4, \ldots, \left\lfloor \frac{1}{\Delta_{\max}} \right\rfloor$ subintervals on the interval $[0, 1]$.

For the (Dλ) model, the estimators are given in closed form in (16) and the likelihood value is easily calculated via Equation (15) and subsequently used for the calculation of the IC. We decide to use the biased estimator $\hat{\sigma}_i^2$: Since we are mainly interested in model selection, we would like to ensure that we work with the optimal value of the log-likelihood when calculating the IC (see also 4.3).
In order to fit the (Pλ) model, we assume that the estimates for $\{\mu_i\}, \{\sigma_i\}$ and $\{\lambda_i\}$ for the (Dλ)-algorithm are already calculated and can be used as an input for the estimation of the (Pλ)-model. As mentioned previously, the estimators for μ_i and σ_i coincide in both models and no further calculation is needed for these parameters. It remains to solve the following minimization problem:

$$\begin{aligned}(\hat{a}, \hat{b}, \hat{c}) &= \underset{a,b,c\in\mathbb{R}}{\arg\min} \left[- \sum_{i=1}^{n} \mathcal{L}_i^P(a, b, c, \mu_i, \sigma_i) \right] \\ &\text{s.t.} \quad a > 0 \text{ and } c > \frac{b^2}{4a} \end{aligned} \tag{26}$$

A reasonable choice of the starting value for the minimization algorithm can be easily obtained by the least-squares fit of the parabola to the $\{\lambda_i\}$ values of the (Dλ) case, which already gives a fairly good approximation of the parabola. In case the initial values obtained by this method do not lie in the admissible set, a change of signs for a or a shift of the parabola may be applied.
Note that the estimation of the (Pλ)-model requires a grid with at least 4 grid points, i.e. 3 subintervals on which $\lambda_1, \lambda_2, \lambda_3$ are estimated using the (Dλ)-model. This ensures that the parabola is well determined. However, as mentioned before, this condition is not restrictive and covers all models on which we would like to run model selection.

4.2.2. ACD model

For both simulation and MLE of ACD models we use the R package `ACDm` written by Markus Belfrage [98]. The model selection analysis for the ACD model follows the Monte Carlo experiment conducted in [99]. We consider model orders $m, q \in \{1, 2\}$ and Table 5 shows the choice of parameters for the simulation.

4.2.3. Numerical results

We use the MSE as a measure for the goodness of fit: Let θ be a generic model parameter to be estimated and $\hat{\theta}$ the corresponding estimator. Given N samples and $\hat{\theta}^{(k)}$, $k = 1, \ldots, N$, the estimates for each sample we calculate the mean squared error to be

$$\mathrm{MSE}(\theta) = \mathbb{E}\left[|\theta - \hat{\theta}|^2\right] = \frac{1}{N}\sum_{k=1}^{N} |\theta - \hat{\theta}^{(k)}|^2. \tag{27}$$

Compound Poisson models. We have to point out first that the distance in Equation (27) has to be understood as a functional distance. To be more precise, we choose the L^2-distance between the true step function intensity and the estimated one:

$$\mathbb{E}\left[|\theta - \hat{\theta}|^2\right] = \mathbb{E}\left[\|\theta - \hat{\theta}\|_{L^2}^2\right] \tag{28}$$

The cases of μ and σ^2 are the easier ones, as we just need to calculate the distance between a step function and a constant: For the step functions with values $\{\mu_i\}$ on the fitting grid $t_1 < t_2 < \ldots < t_n$ Equation (28) can be further written as

$$\begin{aligned}
\mathbb{E}\left[\|\mu - \hat{\mu}\|_{L^2}^2\right] &= \frac{1}{N}\sum_{k=1}^{N} \|\mu - \hat{\mu}^{(k)}\|_{L^2}^2 \\
&= \frac{1}{N}\sum_{k=1}^{N} \int_0^T (\mu(t) - \hat{\mu}^{(k)}(t))^2 \mathrm{d}t \\
&= \frac{1}{N}\sum_{k=1}^{N} \sum_{i=2}^{n} (\mu - \hat{\mu}_i^{(k)})^2 (t_i - t_{i-1}).
\end{aligned} \tag{29}$$

and in the same way for σ^2.

Concerning the intensity function, we have to merge the simulation grid $t_1^s < t_2^s < \ldots < t_m^s$ with the fitting grid $t_1^f < t_2^f < \ldots < t_r^f$. After reordering and relabeling, we can calculate the MSE on the merged grid $t_1 < t_2 < \ldots < t_n$ via

$$\mathbb{E}\left[\|\lambda - \hat{\lambda}\|_{L^2}^2\right] = \frac{1}{N}\sum_{k=1}^{N}\sum_{i=2}^{n}(\lambda_i - \hat{\lambda}_i^{(k)})^2(t_i - t_{i-1}). \tag{30}$$

The numerical results we present here are for $N = 1000$ samples of data simulated from a grid containing 30 subintervals.

Table 6 shows summary statistics of μ and σ^2, where the summary statistics were calculated over the set of fitting grids. The MSE for the μ and σ^2 are comparably small.
For the intensity function λ we plot the MSE against the number of subintervals used for fitting in Figure 12. Starting from a small number of subintervals, the MSE decreases sharply before it reaches its optimum at 30, the true number of subintervals from the simulation. Number of subintervals above 30 give a larger MSE and, in the case of the (Dλ) model, instabilities of over parametrization even lead to an increasing MSE.

Concerning goodness of fit, we can see that the MSE of the (Pλ)-model is consistently smaller than the MSE of the (Dλ)-model. This is to be expected as, by construction of the experiment, the (Pλ)-model is the true model and gives a better fit to the data.
Moreover, we can observe that apart from the optimum at 30 there are "preferred" numbers of subintervals at 10, 20, 45, 60. This is crucial for the explanation of the behavior of model selection as the relationship between goodness of fit and number of subintervals in the region below the optimal number is not monotone.

The size of the MSE can be estimated from the expected fluctuations of the estimator $\hat{\lambda}$. The MSE can be estimated from below by means of the ideal situation when the simulation and fitting grid are identical. Without loss of generality, we assume an equidistant simulation grid with grid size $w = t_i - t_{i-1}$ and rewrite Equation (30):

$$\begin{aligned}\mathbb{E}\left[\|\lambda - \hat{\lambda}\|_{L^2}^2\right] &\geq w\sum_{i=2}^{n}\mathbb{E}\left[(\lambda_i - \hat{\lambda}_i)^2\right] \\ &= w\sum_{i=2}^{n}\mathrm{Var}\left[\hat{\lambda}_i\right] = \frac{1}{w}\sum_{i=2}^{n}\mathrm{Var}\left[N_i\right],\end{aligned} \tag{31}$$

where we have used the definition of the estimator in (16) and the fact that the number of events in an interval of size w is Poisson distributed:

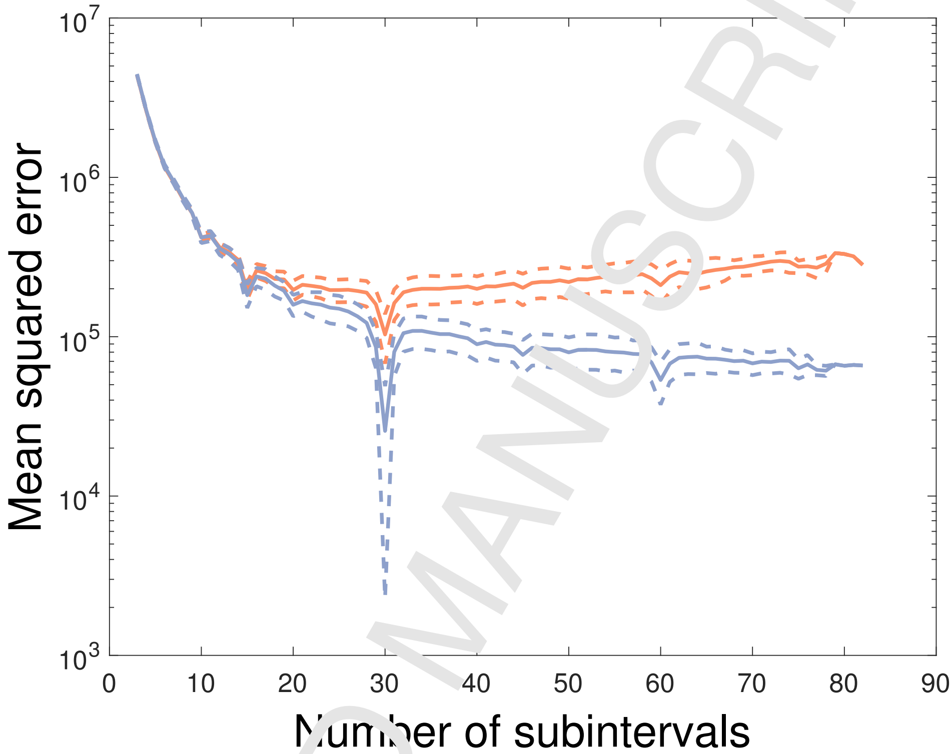

Figure 12: (Color online) Plot of the mean squared error (MSE) of the estimation of the intensity function for the (Dλ)-model (orange lines) and for the (Pλ)-model (blue lines) respectively. The graph shows the MSE together with dashed lines indicating the size of the first standard deviation from the mean as a function of the underlying number of intervals of the fitting grid. The true model for the simulation originally used 30 subintervals. The MSE is calculated as a squared L^2 distance between the estimated and the true intensity function (see also Eq. (30)).

$N_i \sim \text{Poi}(\lambda w)$. We finally get that

$$\mathbb{E}\left[\|\lambda - \hat{\lambda}\|_{L^2}^2\right] \geq \frac{1}{w}\sum_{i=2}^{n} \text{Var}\left[N_i\right]$$
$$= \frac{1}{w}\sum_{i=2}^{n} \lambda_i w \approx \frac{1}{w}\int_0^1 \lambda(t)\,\mathrm{d}t, \tag{32}$$

where we approximate the integral of the step function by the integral of the smooth intensity parametrization in Equation (25). For our numerical example we have $\frac{1}{w} = 30$ and $\lambda_{\min} = 100$ and $\lambda_{\max} = 10000$. An explicit calculation of above integral gives the rough estimate

$$\mathbb{E}\left[\|\lambda - \hat{\lambda}\|_{L^2}^2\right] \gtrsim 30 \cdot 3400 = \mathcal{O}(10^5), \tag{33}$$

which is of about the same order of magnitude observable in Figure 12.

ACD model. In the ACD case we have a simple parameter vector $(\omega, \alpha_1, \ldots,$-$\alpha_m, \beta_1, \ldots, \beta_q) \in \mathbb{R}^{1+m+q}$, Therefore, we can use the formula given in Equation (27) for each scalar valued parameter. The results can be seen in Table 7. The largest sample size ensures that the MSE are comparably low for each model. The largest contribution to the MSE comes from the ω parameter. An even closer look shows that the MSE of the β parameter(s) is of different order depending on the model order q. In the case $q = 1$, the MSE of the β parameter is of the same size as the α parameter(s). However, in the case of $q = 2$, the order of the MSE of the β parameters are significantly larger than the MSE of the α parameters (by a factor of 10 in the ACD$(1, 2)$ case and by a factor of 100 in the ACD$(2, 2)$ case).

4.3. Information criteria and model selection

Starting off from the estimation results in the previous section, we would like to analyse how effective model selection based on information criteria (IC) performs for both the coumpounds Poisson models and the ACD model. As seen in the previous Monte Carlo simulation choosing smaller values of w, i.e. increasing the number of model parameters, gives better fits and the model is able to capture all distributional properties of the quantity of interest. However, a model containing a large number of parameters is likely to be over-fitted. A quantitative method to resolve this trade-off situation is to apply IC. In the following, we will consider three of the most common information criteria:
For a given model fitted to data via MLE let $\mathcal{L}$ be the maximal log-likelihood value, k the number of parameters and T be the sample size of the data set. Then we define:

1. **Akaike's information criterion (AIC)** (see [100])

$$\mathrm{AIC} = -2\mathcal{L} + 2k \tag{34}$$

2. **Bayesian information criterion (BIC)** (see [101])

$$\mathrm{BIC} = -2\mathcal{L} + k\ln(T) \tag{35}$$

3. **Hannan and Quinn information criterion (HQ)** (see [102] and [103])
$$\mathrm{HQ} = -2\mathcal{L} + 2k\ln(\ln(T)) \tag{36}$$

Note that the information criteria under consideration penalize the log-likelihood value for increasing number of parameters k. Among several candidate models, one chooses the model with the smallest IC value. A time grid $t_0 < t_1 < \ldots < t_n$ is given and divides the overall time interval in n subintervals. From Section 4.2.1, we recall that we do not consider $n \in \{1,2\}$. Then the (Dλ)-model has in total $k = 3n$ parameters with $n \in \{3,4,\ldots\}$. This will also be the true number of parameters we expect the IC to choose. In the same way we have for the (Pλ)-model $k = 2n+3$ parameters with $n \in \{3,4,\ldots\}$.

4.3.1. Numerical results

Compound Poisson models. Figures 13, 14 and 15 show box plots of the model selection results of the AIC, BIC and HQ respectively. In each box plot, the orange and blue box plot correspond to the results of the (Dλ)- and (Pλ)-model respectively. The horizontal axis shows the number of subintervals used in the simulation grid. On the vertical axis are the selected number of parameters after the parameter estimation of the (Dλ)- and (Pλ)-models using different discretizations of $[0,1]$. A single box in the box plots extends from the 25th percentile to the 75th percentile and the dot indicates the median. The whiskers have a maximum length of 1.5 times the box length and extend to the outermost point which is not considered as outlier. The crosses indicate outliers.

Below the box plots, bars indicate the ratio of samples which allow model selection under correct specification (blue) and under misspecification (red): In our setting, we speak of model selection under misspecification if the correct model is not contained in the set of selectable models and cannot be chosen by the IC. If this is not the case, i.e. the correct model can potentially be chosen by the IC, we call it model selection under correct specification.

The results for the (Dλ) and (Pλ) model are very similar. Common for all three IC is that for small parameter numbers below 15 the model selection works well: the distributions of the selected orders are concentrated and closely follow the $3n$ or $2n+3$ reference line respectively, where n is the number of subintervals. For very large parameter numbers one can observe that the selected model orders remain distributed around a maximum

model order and stop to follow the linear trend of the reference line. This is rather due to the limitations of our MC setup than the inherent property of the IC: As described in Section 4.2.1, we only work with equidistant grids when applying the model selection procedure. The finest grid which can be used for fitting is determined by the maximal distance $\Delta_{\max}$ between two consecutive points within a sample. On the other hand, $\Delta_{\max}$ is related to the minimal value of λ in the middle of the interval, depending on how small we choose the simulation grid size Δ_{sim}. This means that whenever $\Delta_{\max} > \Delta_{\text{sim}}$, the true model is not contained in the pool of models from which the IC may choose from. In other words, we have a case of model selection under misspecification. The bar plots show that first cases occur at around $n = 20$ and go up to a ratio of about 50% for the finest grid in the analysis.
Another look at Figure 12 hints that the rule "the more parameters, the better the fit" is not entirely true: we can observe that the relation between grid size and MSE is not entirely monotone. This is due to the fact that the fit of the specific model does not only depend on the number of parameters, but also to some extent on the position of the grid. As a consequence, under misspecification, the selected order does not necessarily correspond to the finest available grid size above Δ_{sim}. This might explain the "plateaus" on the model selection results for large parameters.

Between the region of very small and very large parameters the ICs exhibit quite different behaviors according to their intrinsic tendency of under- and overfitting, which will be described in the following:
The AIC tends to overestimate the number of parameters. It allows outliers (in the region of $n \leq 22$) as well as a larger number of cases of the model selection to lie above the reference line (in the region of $n \geq 23$). In contrast, the selected model orders of the BIC and HQ are either on the reference line or strictly below the reference line. In other words BIC and HQ tend to underestimate. Additionally, we can see that for the AIC the boxplot starts to deviate from the reference line starting around $n = 25$ to $n = 27$ and the BIC and HQ deviate earlier around $n = 15$ and $n = 20$ respectively. Especially for $n < 27$ the underestimation in the BIC and HQ case is not attributable to the behaviour of model selection under misspecification, as the ratio of model selection under misspecification is rather low. Based on our results, if the ICs were to be ordered by their parsimonious character, the BIC would be the more parsimonious whereas the AIC the least.
The above observations show that the model selection using any of the three ICs works quite well as long as the true model is actually retrievable. The

AIC tends to overestimate, but the model selection results are closest to the reference line of true parameters compared to the other two ICs.

ACD model. The results of the model selection experiment can be found in Tables 8 to 11. The numbers are success rates in percent of the respective IC to select the correct model from which the simulation data was generated from. The qualitative behaviour of the ICs is not surprisingly similar to the findings for the GARCH model in [99].

A closer look at Table 8 shows that the success rate of the ICs is exceptionally good in the case of ACD$(1, 1)$ data. Even for a small sample size all information criteria are able to detect the correct model order in the majority of cases. The tendency to under-fit works in favour for the BIC and to some extent also for the HQ. For the same reason, the success rates for the AIC are relatively low due to its overfitting property.
A similar behaviour can be observed for ACD$(2, 1)$ in Table 10: Although the IC underestimate the model for smaller sample sizes as a ACD$(1, 1)$ model, they improve for large sample sizes.
In both the ACD$(1, 1)$ and the ACD$(2, 1)$ case, i.e. the cases for $q = 1$, the behaviour of the model selection is acceptable: a reasonably large sample size, which is of the order of a typical intra day trading data sample, ensures a sufficiently large success rate in detecting the correct model. Unfortunately, this cannot be said about the case $q = 2$:
In the first example of ACD$(1, 2)$ data in Table 9, we see that the correct model order is never detected in the majority of cases even for large sample sizes. The best success rates are the ones of the AIC again due to its overfitting tendency. This may be concerning, as this shows that despite the fact that ACD$(1, 2)$ and ACD$(2, 1)$ have the same number of parameters the model selection behaviour is far from comparable.
In comparison, the results for the ACD$(2, 2)$, the most complex model in our experiment, are even more critical: Not only are the IC unable to detect the correct model in most of the cases even with large samples, but the best success rates, again from the AIC, are below 20%.

As mentioned in Section 4.2.3, the cases where model selection fails align with relatively high MSE of the β parameters for $q = 2$: The contribution of the MSE of the ω parameter is not as important, as this parameter is included in all models. However, the increase in MSE when moving from $q = 1$ to $q = 2$ might be one of the factors explaining the discrepancy in model selection between $q = 1$ and $q = 2$. This part of our MC experiment suggests that parameters which are harder to estimate compared to other model parameters (in our case α vs. β parameters or in other words moving average vs. autoregressive parameters in Equation (23)) might also be less likely to be detected by model selection.

5. Discussion

The models analysed in Section 4 are based on the preliminary results presented in [69]. The main idea of that paper was to locally approximate a non-stationary process with a simple normal compound Poisson process. However, many mathematical aspects still need to be clarified. In particular, the choice of the normal compound Poisson process is suggested by the fact that many distributions of positive random variables (the waiting times) can be written as a mixture of exponential distributions. To be more precise, suppose that $\bar{F}_J(u) = \mathbb{P}(J > u)$ is the complementary cumulative distribution function of the positive random variable J. We want to write

$$\bar{F}_J(u) = \int_0^\infty \exp(-\lambda u)\, g(\lambda)\, d\lambda. \tag{37}$$

For instance, from the corollary on page 440, Chapter XIII.4 of Feller [104], we know that the necessary and sufficient condition for a function $\varphi(u)$ to be of the form

$$\varphi(u) = \int_0^\infty \exp(-\lambda u)\, g(\lambda)\, d\lambda$$

when $0 \le g \le C$ is that

$$0 \le \frac{(-x)^n \varphi^{(n)}(x)}{n!} \le \frac{C}{x}; \tag{38}$$

for all $x > 0$. Notice that if $g(\lambda)$ is a continuous probability density function with $g(0)$ equal to some finite non-negative constant, then the condition $0 \le g \le C$ is automatically satisfied. Incidentally, this does not exclude that the representation (37) can be written also when the boundedness hypothesis for g of Feller's corollary are not satisfied.

Similarly, distributions of random variables with support in $\mathbb{R}$ (the log-returns) can be written as a mixture of normal distributions. In particular, the theory of scale mixtures is well-developed [105–108]. Scale mixtures are mixtures of normal distribution with random variance. It turns out that the Laplace [109], the stable family, the Student t family, among others, are scale mixtures. The theory of scale mixtures in [105] is essentially based on the results reported by Feller outlined above and on Bernstein's theorem [110]. Generalizations of the theory to normal variance-mean mixtures do exist [111].

Finally, the local approximation of a non-stationary process with a compound Poisson process naturally follows the evolution of the non-stationary process while activity and volatilty change during the trading day, leading to a satisfactory characterization of the non-stationary behaviour as illustrated in Figure 16 to be compared to Figure 8.

6. Conclusions

In this paper, we addressed two questions. The first one concerns to so-called stylized facts for high-frequency financial data. In particular, do the statistical regularities detected in the past still hold? We cannot give a negative answer to this question. Indeed, we find that some of the scaling properties for financial returns are still approximately satisfied. Most of the studies we refer to concerned a different market (the US NYSE) and were performed several years ago. However, one of the first econophysics papers (if not the first one) concerned returns in the Italian stock exchange (see [112]) and, for this reason, we decided to focus on this market.

The second question is: Is it possible to approximate the non-stationary behavior of intra-day tick-by-tick returns by means of a simple phenomenological stochastic process? We cannot give a negative answer to this question, so far. In Section 4, we present a simple non-homogeneous normal compound Poisson process and we argue that it can approximate empirical data. The cost for simplicity is potential over-fitting as we have to estimate many parameters, but the outcome is a rather accurate representation of the real process. Whether it is possible to rigorously prove convergence of the method outlined in Section 4 is subject to further research and it is outside the scope of the present paper. It is well-known that Lévy processes, namely stochastic processes with stationary and independent increments, can be approximated by compound Poisson processes. The method described in Section 4 can provide a clue for a generalization of such a result to processes with non-stationary and non-independent increments.

Concerning the issue of overfitting, the second part of Section 4 shows that IC are able to detect model orders correctly to some extent when applied to simulated data. It remains to check how well the model selection method performs on empirical data. As a consequence from the numerical results, due to the high variability of model selection in the region of larger numbers of parameters it is not advisable to rely only on the IC based model selection.

It is recommended to combine these with further cross-validation techniques. A similar conclusion holds for the ACD model, as model selection using IC is adversely affected by differing MLE quality for different model orders.

Acknowledgments

This work was partially supported by MIUR PRIN 2009 grant *The growth of firms and countries: distributional properties and economic determinants - Finitary and non-finitary probabilistic models in economics 2009H8WPX5_002* and by an SDF fund provided by the University of Sussex.

References

1. C. Goodhart, M. O'Hara, High-frequency data in financial markets: Issues and applications, J. of Empir. Financ. 4 (1997) 73–114.

2. M. O'Hara, Making market microstructure matter, Financ. Manage. 28 (1999) 83–90.

3. A. Madhavan, Market microstructure: A survey, Journal of Financial Markets 3 (2000) 205–258.

4. E. Scalas, R. Gorenflo, F. Mainardi, Fractional calculus and continuous-time finance, Physica A 284 (2000) 376–384.

5. F. Mainardi, M. Raberto, R. Gorenflo, E. Scalas, Fractional calculus and continuous-time finance II: the waiting-time distribution, Physica A 287 (3-4) (2000) 468–481.

6. M. Dacorogna, R. Gençay, U. Müller, R. B. Olsen, O. Pictet, An Introduction to High Frequency Finance, Academic Press, 2001.

7. M. Raberto, S. Cincotti, S. M. Focardi, M. Marchesi, Agent-based simulation of a financial market, Physica A 219 (2001) 319–327.

8. S. Cincotti, S. M. Focardi, M. Marchesi, M. Raberto, Who wins? Study of long-run trader survival in an artificial stock market, Physica A 324 (1-2) (2003) 227–233.

9. H. Luckock, A steady-state model of the continuous double auction, Quant. Finance 3 (2003) 385–404.

10. E. Scalas, R. Gorenflo, H. Luckock, F. Mainardi, M. Mantelli, M. Raberto, Anomalous waiting times in high-frequency financial data, Quant. Finance 4 (2004) 1–8.

11. S. Pastore, L. Ponta, S. Cincotti, Heterogeneous information-based artificial stock market, New J. Phys. 12 (2010) 053035.

12. L. Ponta, E. Scalas, M. Raberto, S. Cincotti, Statistical analysis and agent-based microstructure modeling of high-frequency financial trading, IEEE Journal of Selected Topics in Signal Processing 6 (2012) 381–387.

13. L. Ponta, S. Cincotti, Traders' networks of interactions and structural properties of financial markets: An agent-based approach, Complexity 2018.

14. L. Ponta, S. Pastore, S. Cincotti, Static and dynamic factors in an information-based multi-asset artificial stock market, Physica A: Statistical Mechanics and its Applications 492 (2018) 814–823.

15. B. Mandelbrot, The variation of certain speculative prices, J. Business 36 (1963) 394–419.

16. B. Mandelbrot, Fractals and Scaling in Finance, Berlin: Springer, 1997.

17. U. Müller, M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz, C. Morgenegg, Statistical study of foreign exchange rates, J. Bank. Financ. 14 (1990) 1189–1208.

18. R. N. Mantegna, H. E. Stanley, Scaling behavior in the dynamics of an economic index, Nature 376 (6535) (1995) 46–49.

19. P. Gopikrishnan, V. Plerou, X. Gabaix, H. E. Stanley, Statistical properties of share volume traded in financial markets, Phys. Rev. E 62 (2000) 4493–4496.

20. N. Hautsch (Ed.), Econometrics of Financial High-Frequency Data, Springer, Berlin, 2012.

21. V. Gontis, B. Kaulakys, Long-range memory model of trading activity and volatility, Journal of Statistical Mechanics P10016 (2006) 1–11.

22. V. Gontis, B. Kaulakys, Modeling long-range memory trading activity by stochastic differential equations, Physica A 382 (2007) 114–120.

23. V. Gontis, J. Ruseckas, A. Kononovicius, A long-range memory stochastic model of the return in financial markets, Physica A 389 (2010) 100–106.

24. V. Gontis, A. Kononovicius, Nonlinear stochastic model of return matching to the data of new york and vilnius stock exchanges, Dynamics of Socio-Economic Systems 2 (2011) 101–109.

25. B. Kaulakys, M. Alaburda, V. Gontis, Point processes modeling of time series exhibiting power-law statistics, AIP Conference Proceedings 922 (2007) 535– 538.

26. B. Kaulakys, M. Alaburda, V. Gontis, T. Meskauskas, J. Ruseckas, Modeling of flows with power-law spectral densities and power-law distributions of flow intensities, in: A. Schadschneider (Ed.), Traffic and granular flow, Vol. 5, Springer, 2007, pp. 587–594.

27. B. Kaulakys, M. Alaburda, V. Gontis, Long-range stochastic point processes with the power law statistics, in: M. H. M. Janzura (Ed.), Proceeding of Prague Conference, Matfyzpress, Charles University in Prague, Prague, 2006, pp. 364–373.

28. B. Kaulakys, M. Alaburda, V. Gontis, T. Meskauskas, Multifractality of the multiplicative autogressive point processes, in: M. M. Novak (Ed.), Complexus Mundi: Emergent Patterns in Nature, World Scientific, 2006, pp. 277–286.

29. D. Y. Kenett, E. Ben-Jacob, H. E. Stanley, G. Gur-Gershgoren, How high frequency trading affects a market index, Scientific Reports 3 (2013) 2110.

30. Z. Zheng, Z. Qiao, T. Takaishi, H. E. Stanley, B. Li, Realized volatility and absolute return volatility: A comparison indicating market risk, PLOS ONE 9 (7) (2014) e102940.

31. F. Botta, H. S. Moat, H. E. Stanley, T. Preis, Quantifying stock return distributions in financial markets, PLOS ONE 10 (9) (2015) e0135600.

32. R. F. Engle, J. R. Russell, Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model, J. Empir. Financ. 4 (1997) 187–212.

33. R. F. Engle, J. R. Russell, Autoregressive conditional duration: A new model for irregularly spaced transaction data, Econometrica 66 (1998) 1127–1162.

34. L. Bauwens, P. Giot, The logarithmic acd model: An application to the bid-ask quote process of three nyse stocks, Ann. Econ. Stat. 60 (2000) 117–149.

35. A. W. Lo, A. C. MacKinlay, J. Zhang, Econometric models of limit-order executions, J. Financ. Econ. 65 (2002) 31–71.

36. R. Cont, J.-P. Bouchaud, Herd behavior and aggregate fluctuations in financial markets, Macroecon. Dyn. 4 (2) (2000) 170–196.

37. D. Chowdhury, D. Stauffer, A generalized spin model of financial markets, Eur. Phys. J. B 8 (1999) 477.

38. W. Hardle, A. Kirman, Neoclassical demand - a model-free examination of price-quantity relations in the marseilles fish market, J. Econometrics 67 (1995) 227–257.

39. M. Levy, H. Levy, S. Solomon, Microscopic simulation of the stock market: The effect of microscopic diversity, J. Phys. I France 5 (1995) 1087–1107.

40. T. Lux, M. Marchesi, Scaling and criticality in a stochastic multi-agent model of a financial market, Nature 397 (6718) (1999) 498–500.

41. D. Stauffer, D. Sornette, Self-organized percolation model for stock market fluctuations, Physica A 271 (1999) 496–506.

42. M. Youssefmir, B. A. Huberman, Clustered volatility in multiagent dynamics, J. Econ. Behav. Organ. 32 (1) (1997) 101–118.

43. A. Hawkes, Spectra of some self-exciting and mutually exciting point processes, Biometrika 58 (1971) 83–90.

44. I. Muni Toke, F. Pomponio, Modelling trades-through in a limit order book using Hawkes processes, Economics: The Open-Access, Open-Assessment E-Journal 6 (2012) 2012–22.

45. V. Filimonov, D. Sornette, Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data, Quant. Finance 15 (2015) 1293–1314.

46. M. E. Blume, A. C. Mackinlay, B. Terker, Order imbalances and stock price movements on october 19 and 20, 1987, J. Finance 44 (1989) 827–848.

47. K. Chan, W.-M. Fong, Trade size, order imbalance, and the volatility-volume relation, J. Financ. Econ. 57 (2000) 247–273.

48. H. R. Stoll, R. E. Whaley, Stock market structure and volatility, Rev. Financ. Stud. 3 (1990) 37–71.

49. S. Hauser, B. Lauterbach, The impact of minimum trading units on stock value and price volatility, J. Financ. Quant. Anal. 38 (2003) 575–589.

50. T. Chordia, R. Roll, A. Subrahmanyam, Order imbalance, liquidity, and market returns, J. Financ. Econ. 65 (2002) 111–130.

51. A. Ponzi, F. Lillo, R. N. Mantegna, Market reaction to a bid-ask spread change: A power-law relaxation dynamics, Phys. Rev. E 80 (2009) 016112.

52. A. Svorencik, F. Slanina, Interacting gaps model, dynamics of order book, and stock-market fluctuations, Eur. Phys. J. B 57 (2007) 453–462.

53. M. Wyart, J.-P. Bouchaud, J. Kockelkoren, M. Potters, M. Vettorazzo, Relation between bid-ask spread, impact and volatility in order-driven markets, Quant. Finance 8 (2008) 41–57.

54. E. Moro, J. Vicente, L. G. Moyano, A. Gerig, J. D. Farmer, G. Vaglica, F. Lillo, R. N. Mantegna, Market impact and trading profile of hidden orders in stock markets, Phys. Rev. E 80 (2009) 066102.

55. J. Perelló, J. Masoliver, A. Kasprzak, R. Kutner, Model for interevent times with long tails and multifractality in human communications: An application to financial trading, Phys. Rev. E 78 (2008) 036108.

56. T. Preis, J. J. Schneider, H. E. Stanley, Switching processes in financial markets, Proc. Natl. Acad. Sci. U.S.A. 108 (2011) 7674–7678.

57. M. Kumaresan, N. Krejic, A model for optimal execution of atomic orders, Comput. Optim. Appl. 46 (2010) 369–389.

58. A. Zaccaria, M. Cristelli, V. Alfi, F. Ciulla, L. Pietronero, Asymmetric statistics of order books: The role of discreteness and evidence for strategic order placement, Phys. Rev. E 81 (2010) 066101.

59. M. Lim, R. Coggins, The immediate price impact of trades on the australian stock exchange, Quant. Finance 5 (2005) 365–377.

60. P. Weber, B. Rosenow, Order book approach to price impact, Quant. Finance 5 (2005) 357–364.

61. J.-P. Bouchaud, The subtle nature of financial random walks, Chaos 15 (2005) 026104.

62. G. Bonanno, F. Lillo, R. Mantegna, Dynamics of the number of trades of financial securities, Physica A 280 (2000) 136–141.

63. V. Plerou, P. Gopikrishnan, L. A. Nunes Amaral, X. Gabaix, H. E. Stanley, Economic fluctuations and anomalous diffusion, Phys. Rev. E 62 (2000) 3023–3026.

64. E. Scalas, The application of continuous-time random walks in finance and economics, Physica A 362 (2006) 225–239.

65. V. Gontis, B. Kaulakys, J. Ruseckas, Trading activity as driven Poisson process: Comparison with empirical data, Physica A 387 (2008) 3891–3896.

66. C. Tsallis, Inter-occurrence times and universal laws in finance, earthquakes and genomes, Chaos, Solitons & Fractals 88 (2016) 254–266.

67. R. L. Hudson, B. B. Mandelbrot, The (Mis)Behaviour of Markets, Profile Business, 2010.

68. S. Vrobel, Fractal time why a watched kettle never boils, in: B. J. West (Ed.), Studies Of Nonlinear Phenomena In Life Science, World Scientific, Imperial College Press, 2011.

69. E. Scalas, Mixtures of compound Poisson processes as models of tick-by-tick financial data, Chaos Soliton. Frac. 34 (2007) 33–40.

70. M. Politi, E. Scalas, Fitting the empirical distribution of intertrade durations, Physica A 387 (2008) 2025–2034.

71. M. Raberto, E. Scalas, F. Mainardi, Waiting times and returns in high-frequency financial data: an empirical study, Physica A 314 (1-4) (2002) 749–755.

72. J. Masoliver, M. Montero, G. H. Weiss, Continuous-time random-walk model for financial distributions, Phys. Rev. E 67 (2003) 021112.

73. P. C. Ivanov, A. Yuen, B. Podobnik, Y. Lee, Common scaling patterns in intertrade times of U.S. stocks, Phys. Rev. E 69 (2004) 056107.

74. A. Kasprzak, R. Kutner, J. Perelló, J. Masoliver, Higher-order phase transitions on financial markets, The European Physical Journal B 76 (4) (2010) 513–527.

75. R. Kutner, J. Masoliver, The continuous time random walk, still trendy: fifty-year history, state of art and outlook, The European Physical Journal B 90 (3) (2017) 50.

76. M. Goldstein, S. Morris, G. Yen, Problems with fitting to the power-law distribution, Eur. Phys. J. B. 41 (2004) 255–258.

77. P. Embrechts, C. Klüppelberg, T. Mikosch, Modelling Extremal Events for Insurance and Finance, Springer, Berlin, 1997.

78. W. Press, B. Flannery, S. Teukolsky, Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, 1992.

79. F. Mainardi, R. Gorenflo, E. Scalas, A fractional generalization of the Poisson process, Vietnam J. Math. 32 (2004) 53–64.

80. R. F. Engle, J. R. Russell, Forecasting transaction rates: the autoregressive conditional duration model, NBER Working paper series (1994) 4966.

81. M. Denys, T. Gubiec, R. Kutner, M. Jagielski, H. E. Stanley, Universality of market superstatistics, Physical Review E 94 (4) (2016) 042305.

82. S. Camargo, S. M. Duarte Queirós, C. Anteneodo, Nonparametric segmentation of nonstationary time series, Phys. Rev. E 84 (2011) 046702.

83. S. Camargo, S. M. Duarte Queirós, C. Anteneodo, Bridging stylized facts in finance and data non-stationarities, Eur. Phys. J. B 86 (2013) 159.

84. T. Gubiec, M. Wiliński, Intra-day variability of the stock market activity versus stationarity of the financial time series, Physica A: Statistical Mechanics and its Applications 432 (2015) 216–221.

85. M. B. Graczyk, S. M. Duarte Queirós, Intraday seasonalities and nonstationarity of trading volume in financial markets: Individual and cross-sectional features, PLOS ONE 11 (11) (2016) e0165057.

86. W. K. Bertram, A threshold model for Australian stock exchange equities, Physica A 346 (2005) 561–576.

87. G. Livan, J. Inoue, E. Scalas, On the non-stationarity of financial time series: Impact on optimal portfolio selection, J. Stat. Mech. (2012) P07025.

88. FTSEMIB, Methodology for the management of the FTSE MIB index, Version 2.0, Borsa Italiana, London Stock Exchange Group, London, U.K. (2011).

89. S. Golia, Long memory effects in ultra- high frequency data, Quaderni di Statistica 3 (2001) 43–52.

90. L. Sabatelli, S. Keating, J. Dudley, P. Richmond, Waiting time distribution in financial markets, Eur. Phys. J. B 27 (2002) 273–275.

91. H. Takayasu (Ed.), Empirical Science of Financial Fluctuations: The Advent of Econophysics, Springer, Tokyo, 2002.

92. C. Marinelli, S. T. Rachev, R. Roll, Subordinated exchange rate models: Evidence for heavy tailed distributions and long-range dependence, Math. Comput. Modell. 34 (2001) 955–1001.

93. D. Stauffer, H. E. Stanley, From Newton to Mandelbrot: A Primer in Theoretical Physics, Springer, Berlin, 1995.

94. A. Bunde, S. Havlin (Eds.), Fractals in Science, Springer, Berlin, 1994.

95. Z. Eisler, J. Kertész, Size matters: some stylized facts of the stock market revisited, Eur. Phys. J. B.

96. W. K. Bertram, An empirical investigation of Australian stock exchange data, Physica A 341 (2004) 533–546.

97. D. L. Snyder, M. I. Miller, Random Point Processes in Time and Space, Springer, 1991, Ch. 4.4.

98. M. Belfrage, ACDm: Tools for Autoregressive Conditional Duration Models, r package version 1.0.4 (2016).
URL `https://CRAN.R-project.org/package=ACDm`

99. F. Javed, P. Mantalos, GARCH-type models and performance of information criteria, Comm. Statist. Simulation Comput. 42 (8) (2013) 1917–1933.

100. H. Akaike, Information theory and an extension of the maximum likelihood principle, in: Second International Symposium on Information Theory (Tsahkadsor, 1971), Akadémiai Kiadó, Budapest, 1973, pp. 267–281.

101. G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (2) (1978) 461–464.

102. E. J. Hannan, B. G. Quinn, The determination of the order of an autoregression, J. Roy. Statist. Soc. Ser. B 41 (2) (1979) 190–195.

103. E. J. Hannan, The estimation of the order of an ARMA process, Ann. Statist. 8 (5) (1980) 1071–1081.

104. W. Feller, An Introduction to Probability Theory and its Applications, John Wiley and Sons, 1971.

105. D. Andrews, M. C.L., Scale mixtures of normal distributions, Journal of the Royal Statistical Society. Series B 36 (1) (1974) 99–102.

106. B. Efron, R. A. Olshen, How broad is the class of normal scale mixtures?, The Annals of Statistics 6 (5) (1978) 1159–1164.

107. M. West, Outlier models and prior distributions in bayesian linear regression, Journal of the Royal Statistical Society. Series B 46 (3) (1984) 431–439.

108. M. West, On scale mixtures of normal distributions, Biometrika 74 (3) (1987) 646–648.

109. A. Manas, The Laplace illusion, Physica A 391 (2012) 3963–3970.

110. S. Bernstein, Acta Mathematica 52 (1928) 1

111. O. Barndorff-Nielsen, J. Kent, M. Sørensen, Normal variance-mean mixtures and z distributions, International Statistical Review/Revue Internationale de Statistique 50 (2) (1982) 145–149.

112. R. N. Mantegna, Lévy walks and enhanced diffusion in Milan stock exchange, Physica A 179 (1991) 232–242.

Table 1: Descriptive statistics for the waiting times [a]

Asset	mean	std	α	β	AD	Lillie
A2A	32.49	39.04	0.053	0.865	106	0.068
STS	34.07	43.68	0.061	0.818	122	0.083
ATL	24.42	32.48	0.088	0.792	263	[illegible]
AGL	33.20	41.87	0.059	0.830	145	[illegible]
AZM	34.67	42.35	0.052	0.853	116	0.074
BP	9.54	12.80	0.189	0.786	1158	0.134
BMPS	17.21	23.96	0.130	0.761	401	0.107
PMI	19.95	27.26	0.111	0.773	293	[illegible]
BUL	24.87	37.02	0.116	0.717	326	0.123
BZU	22.62	33.71	0.125	0.716	[illegible]	[illegible]
CPR	33.77	42.42	0.058	0.833	[illegible]	0.092
DIA	30.21	39.91	0.073	0.797	155	0.091
ENEL	9.19	11.60	0.173	0.829	[illegible]	0.123
EGPW	21.16	29.31	0.110	0.764	[illegible]	0.094
ENI	8.71	12.21	0.221	[illegible]	[illegible]	0.148
EXO	22.72	31.16	0.101	0.771	228	0.094
F	7.94	11.29	0.243	0.747	1936	0.158
FI	12.80	18.77	0.182	[illegible]	833	0.132
FNC	20.86	26.98	0.093	0.812	234	0.089
FSA	23.70	35.15	0.120	0.719	309	0.118
G	11.10	14.79	0.165	0.792	759	0.119
IPG	32.26	41.41	[illegible]	0.818	157	0.085
ISP	7.96	11.30	[illegible]	0.748	1930	0.158
LTO	33.22	[illegible]	0.062	0.819	117	0.082
LUX	23.28	31.52	[illegible]	0.780	231	0.096
MS	20.12	[illegible]	0.114	0.763	350	0.107
MB	17.40	24.03	0.126	0.767	403	0.108
MED	31.66	[illegible]	0.060	0.837	126	0.077
PLT	[illegible]	[illegible]	0.119	0.749	322	0.104
PC	22.78	30.45	0.094	0.789	221	0.092
PRY	19.48	[illegible]	0.126	0.743	390	0.113
SPM	[illegible]	17.88	0.219	0.691	1185	0.150
SRG	[illegible]	32.77	0.086	0.796	208	0.091
STM	12.22	17.26	0.174	0.751	750	0.124
TIT	[illegible]	20.52	0.198	0.692	972	0.146
TEN	[illegible]	24.98	0.137	0.743	395	0.110
TRN	28.12	35.52	0.068	0.829	148	0.080
TOD	31.31	40.71	0.068	0.808	114	0.081
UBI	20.58	27.30	0.100	0.794	272	0.096
UCG	3.85	4.94	0.364	0.817	8640	0.223
Index	1.66	1.26	–	–	Inf	0.365

Table 2: Descriptive statistics for the trade-by-trade log-returns [illegible]. (*) On March 7th, 2011, the French firm LVMH launched a takeover offer (OPA - *Offerta Pubblica d'Acquisto* in Italian) to buy Bulgari shares at 12.25 euros. On that day, the share price jumped from below 8 euros to more than 12 euros.

Assets	mean $\times 10^{-7}$	variance $\times 10^{-7}$	skewness $\times 10^{-2}$	kurtosis
A2A	29.15	5.24	[illegible].36	5.22
STS	−14.43	6.76	−7.[illegible]1	11.50
ATL	1.59	2.09	[illegible]4.62	19.64
AGL	−36.50	6.09	114.[illegible]	43.47
AZM	−3.29	8.03	−2[illegible].90	14.14
BP	−4.53	4.55	−1.69	10.69
BMPS	24.93	4.79	21.71	24.34
PMI	6.87	5.55	−23.73	41.72
BUL (*)	−3.75	4.37	−295.68	154.69
BZU	61.92	7.41	−99.04	35.92
CPR	2.35	3.73	11.04	8.13
DIA	−40.04	4.42	−49.99	29.17
ENEL	6.21	1.38	140.10	76.06
EGPW	38.81	3.64	3.43	7.31
ENI	7.86	1.4[illegible]	59.89	21.01
EXO	11.98	4.82	−5.45	8.06
F	−3.55	2.81	−45.05	21.76
FI	14.33	[illegible].68	−39.37	18.14
FNC	0.50	3.2[illegible]	28.01	13.01
FSA	84.68	10.3[illegible]	−163.51	180.64
G	5.03	2.[illegible]9	−100.65	44.97
IPG	80.6[illegible]	[illegible].04	−45.81	22.68
ISP	1.99	3.45	−62.87	43.12
LTO	6[illegible].82	9.28	−171.44	62.62
LUX	25.[illegible]	2.67	30.48	24.43
MS	5.76	2.86	−22.98	19.38
MB	[illegible]7.29	4.18	1.66	9.67
MED	[illegible]0.25	7.64	−43.78	18.78
PLT	[illegible]9.76	5.30	49.56	14.43
PC	47.93	5.41	3.44	10.75
PRY	[illegible]1.54	4.02	257.09	92.76
SPM	5.72	1.50	−9.12	32.75
SRG	12.09	2.41	79.03	54.87
STM	15.69	2.56	−39.64	36.78
TIT	8.33	3.20	−22.22	8.92
TEN	0.34	2.61	−112.99	135.05
TRN	26.67	2.42	3.54	6.03
TOD	28.73	6.95	158.96	86.49
UBI	−1.76	4.99	−67.53	25.23
UCG	3.44	1.29	−12.56	57.51
Index	1.10	0.03	2	8.54

Table 3: Kolmogorov-Smirnov test. The null hypothesis of empirical data coming from an identical distribution is rejected in the comparisons of $\Delta t = 3s$ and $\Delta t = 5s$, $\Delta t = 3s$ and $\Delta t = 10s$ and $\Delta t = 3s$ and $\Delta t = 30s$.

Δt	3s	5s	10s	30s	300s
3s	0.000	0.010	0.014	0.014	0.023
5s	0.010	0.000	0.008	0.010	0.022
10s	0.014	0.008	0.000	0.008	0.017
30s	0.014	0.010	0.008	0.000	0.018
300s	0.023	0.022	0.017	0.018	0.000

Table 4: Kolmogorov-Smirnov test. The null hypothesis of simulated data coming from an identical distribution is always rejected.

Δt	3s	5s	10s	30s	300s
3s	0.000	0.019	0.031	0.036	0.035
5s	0.019	0.000	0.012	0.018	0.018
10s	0.031	0.012	0.000	0.007	0.016
30s	0.036	0.018	0.007	0.000	0.019
300s	0.035	0.018	0.016	0.019	0.000

Table 5: Parameter settings for the simulation of ACD data

	ω	α_1	α_2	β_1	β_2
ACD(1,1)	1	0.080	–	0.85	–
ACD(1,2)	1	0.1	–	0.45	0.4
ACD(2,1)	1	0.15	0.15	0.65	–
ACD(2,2)	1	0.1	0.1	0.42	0.35

Table 6: Table of summary statistics of the MSE of the parameters μ and σ^2 of the compound Poisson type model. The analysis is based on 1000 samples generated from a simulation grid containing 30 subintervals.

	mean	min	max	std
μ	0.0545	0.0026	0.1049	0.0212
σ^2	0.1038	0.0049	0.1757	0.0439

Table 7: Results of the MSE calculations for the ACD model

		MSE(ω)	MSE(α_1)	MSE(α_2)	MSE(β_1)	MSE(β_2)
ACD(1,1)	T=250	3.7508	0.0023	–	0.0231	–
	T=500	1.8887	0.0010	–	0.0108	–
	T=1000	0.3591	0.0005	–	0.0025	–
	T=2000	0.1245	0.0002	–	0.0010	–
ACD(1,2)	T=250	14.5255	0.0036	–	0.4748	0.4282
	T=500	3.7468	0.0019	–	0.3039	0.2681
	T=1000	0.6259	0.001[illegible]	–	0.1869	0.1606
	T=2000	0.1905	0.0005	–	0.0809	0.0681
ACD(2,1)	T=250	0.8491	0.0063	0.0108	0.0130	–
	T=500	0.2664	0.0032	0.0050	0.0053	–
	T=1000	0.0916	0.0014	0.0026	0.0023	–
	T=2000	0.0418	0.0007	0.0012	0.0011	–
ACD(2,2)	T=250	6.4135	0.0067	0.0102	0.3165	0.2445
	T=500	1.1077	0.0032	0.0061	0.2722	0.2031
	T=1000	0.37[illegible]	0.0014	0.0041	0.2086	0.1526
	T=2000	0.1512	0.0006	0.0026	0.1612	0.1181

Table 8: Model selection results based on ACD(1,1) data samples: Given 1000 samples of size $T \in \{250, 500, 1000, 2000\}$ each column gives the percentage of cases in which the different IC selected the model ACD(1,1), ACD(1,2), ACD(2,1) and ACD(2,2) respectively. The bold numbers give the largest percentage per row.

		ACD(1,1)	ACD(1,2)	ACD(2,1)	ACD(2,2)
T=250	AIC	**58.7**	23.6	9.9	7.8
	BIC	**90.2**	7	2.1	0.7
	HQ	**77.9**	14.6	4.8	2.7
T=500	AIC	**62.9**	20.4	10.9	5.8
	BIC	**93.6**	4.7	1.6	0.1
	HQ	**82.6**	11.5	4.9	1
T=1000	AIC	**67.5**	16.4	11	5.1
	BIC	**97.4**	1.8	0.8	0
	HQ	**87.2**	7.5	4.8	0.5
T=2000	AIC	**71.3**	13.1	9.7	5.9
	BIC	**97.7**	1.6	0.6	0.1
	HQ	**91.5**	4.4	3	1.1

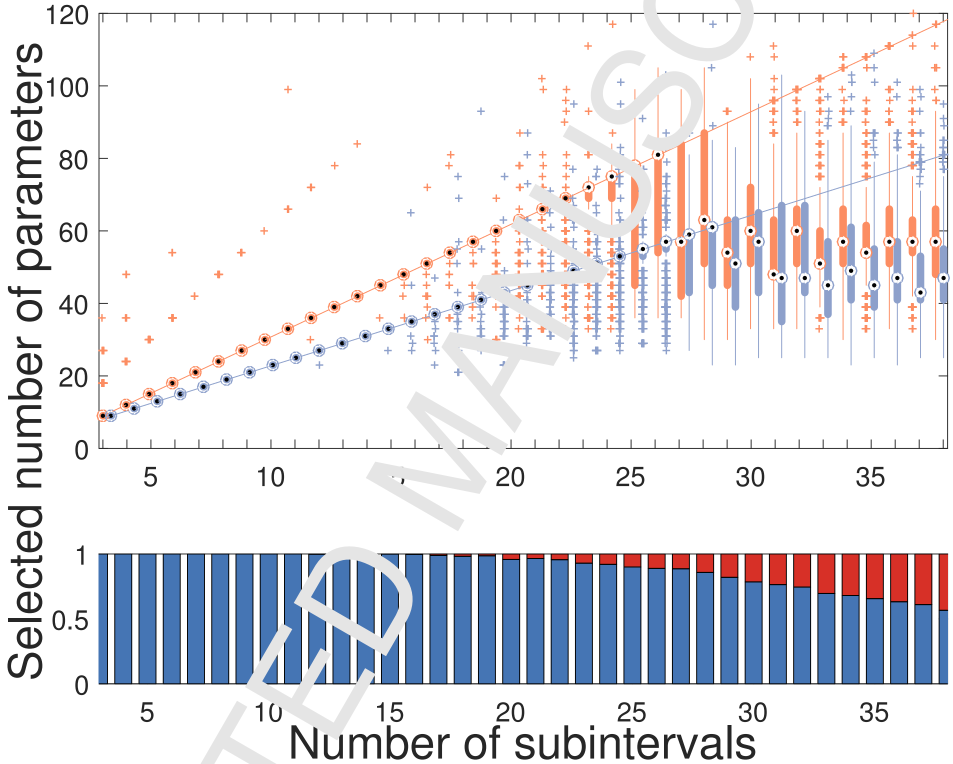

Figure 13: (Color online) The lower plot shows the ratio of samples which allow the true model to be among the set of models from which the IC may choose from, in other words there is no misspecification (blue areas). This ratio decreases and for finer discretization there are more cases of model selection under misspecification (red areas). The sum of blue and red areas is 100%.
The upper plot shows that the model selection using the AIC for the (Dλ)-model (orange box plot) closely follows the reference line indicating $3n$ (n = number of subintervals) for small n before deviating for larger n. The same holds for the (Pλ)-model (blue box plot) and its corresponding reference line $2n + 1$. The number of subintervals for which both box plots deviate from their respective reference lines is around $n = 25$ to $n = 27$. In the region $n < 15$, there are several outliers which are almost all overestimates.

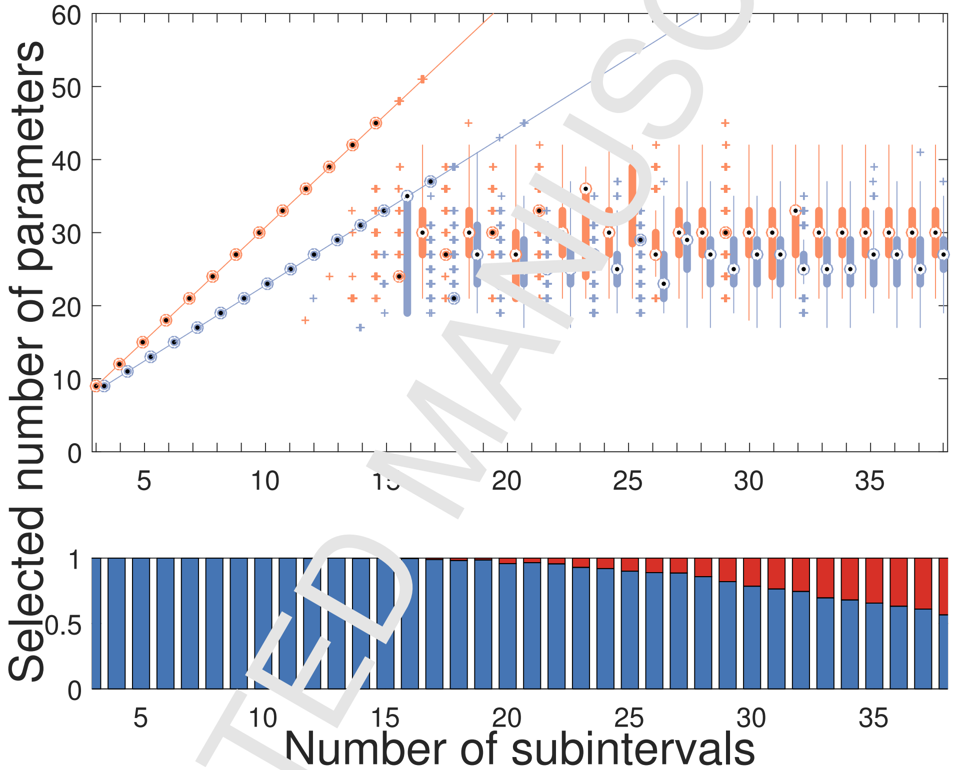

Figure 14: (Color online) The lower plot shows the ratio of samples which allow the true model to be among the set of models from which the IC may choose from, in other words there is no misspecification (blue areas). This ratio decreases and for finer discretization there are more cases of model selection under misspecification (red areas). The sum of blue and red areas is 100%.
The upper plot shows that the model selection using the BIC for the (Dλ)-model (orange box plot) closely follows the reference line indicating $3n$ (n = number of subintervals) for small n before deviating for larger n. The same holds for the (Pλ)-model (blue box plots) and its corresponding reference line $2n+1$. The number of subintervals for which both box plots deviate from their respective reference lines is around $n = 15$ to $n = 17$.

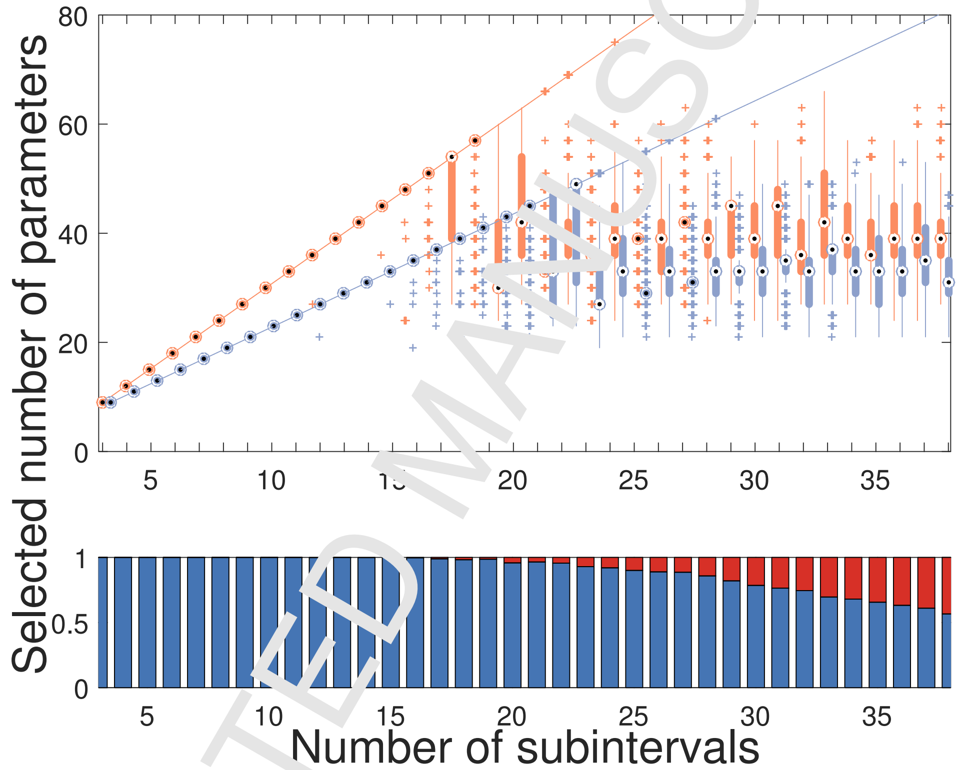

Figure 15: (Color online) The lower plot shows the ratio of samples which allow the true model to be among the set of models from which the IC may choose from, in other words there is no misspecification (blue areas). This ratio decreases and for finer discretization there are more cases of model selection under misspecification (red areas). The sum of blue and red areas is 100%.
The upper plot shows that the model selection using the HQ for the (Dλ)-model (orange box plot) closely follows the reference line indicating $3n$ (n = number of subintervals) for small n before deviating for larger n. The same holds for the (Pλ)-model (blue box plots) and its corresponding reference line $2n + 1$. The number of subintervals for which both box plots deviate from their respective reference lines is around $n = 18$ to $n = 20$.

Table 9: Model selection results based on ACD(1,2) data samples: Given 1000 samples of size $T \in \{250, 500, 1000, 2000\}$ each column gives the percentage of cases in which the different IC selected the models ACD(1,1), ACD(1,2), ACD(2,1) and ACD(2,2) respectively. The bold numbers give the largest percentage per row.

		ACD(1,1)	ACD(1,2)	ACD(2,1)	ACD(2,2)
T=250	AIC	**58.6**	24.7	9.6	7.1
	BIC	**91.5**	6.5	1.3	0.7
	HQ	**78.6**	14.8	3.7	2.9
T=500	AIC	**60.6**	25.1	10.3	4
	BIC	**94.7**	4.3	0.7	0.3
	HQ	**81.2**	13.5	4.5	0.8
T=1000	AIC	**52.7**	27.8	15.2	4.3
	BIC	**92.6**	5.1	2.3	0
	HQ	**76**	14.7	8.8	0.5
T=2000	AIC	**41.5**	35.6	18	4.9
	BIC	**88.4**	6.7	4.9	0
	HQ	**67.6**	20.4	11.6	0.4

Table 10: Model selection results based on ACD(2,1) data samples: Given 1000 samples of size $T \in \{250, 500, 1000, 2000\}$ each column gives the percentage of cases in which the different IC selected the models ACD(1,1), ACD(1,2), ACD(2,1) and ACD(2,2) respectively. The bold numbers give the largest percentage per row.

		ACD(1,1)	ACD(1,2)	ACD(2,1)	ACD(2,2)
T=250	AIC	**36.2**	20.9	31.8	11.1
	BIC	**73.7**	8.9	16.8	0.6
	HQ	**52.4**	16.3	28.1	3.2
T=500	AIC	19.1	20.7	**50**	10.2
	BIC	**59.9**	10.5	29	0.6
	HQ	36.5	16.4	**43.8**	3.3
T=1000	AIC	7.4	16.7	**64.8**	11.1
	BIC	35.6	11.9	**52.1**	0.4
	HQ	17.1	15.7	**63.7**	3.5
T=2000	AIC	1.2	12.7	**74.2**	11.9
	BIC	6.8	12.9	**80.1**	0.2
	HQ	2.2	14.2	**81.6**	2

Table 11: Model selection results based on ACD(2,2) data samples: Given 1000 samples of size $T \in \{250, 500, 1000, 2000\}$ each column gives the percentage of cases in which the different IC selected the models ACD(1,1), ACD(1,2), ACD(2,1) and ACD(2,2) respectively. The bold numbers give the largest percentage per row.

		ACD(1,1)	ACD(1,2)	ACD(2,1)	ACD(2,2)
T=250	AIC	**56.7**	15.3	18.8	8.7
	BIC	**89.7**	5.3	4.5	0.5
	HQ	**74**	11.5	11.7	2.8
T=500	AIC	**57.2**	13.6	19.1	10.1
	BIC	**92.1**	2.9	4.6	0.4
	HQ	**78.4**	8	11.4	2.2
T=1000	AIC	**48.4**	13.1	23.4	15.1
	BIC	**91.5**	2.7	5.7	0.1
	HQ	**74**	6.9	16.1	3
T=2000	AIC	34.2	9.7	**37.2**	18.9
	BIC	**86.1**	1.8	11.5	0.6
	HQ	**59.7**	6.8	26.5	7

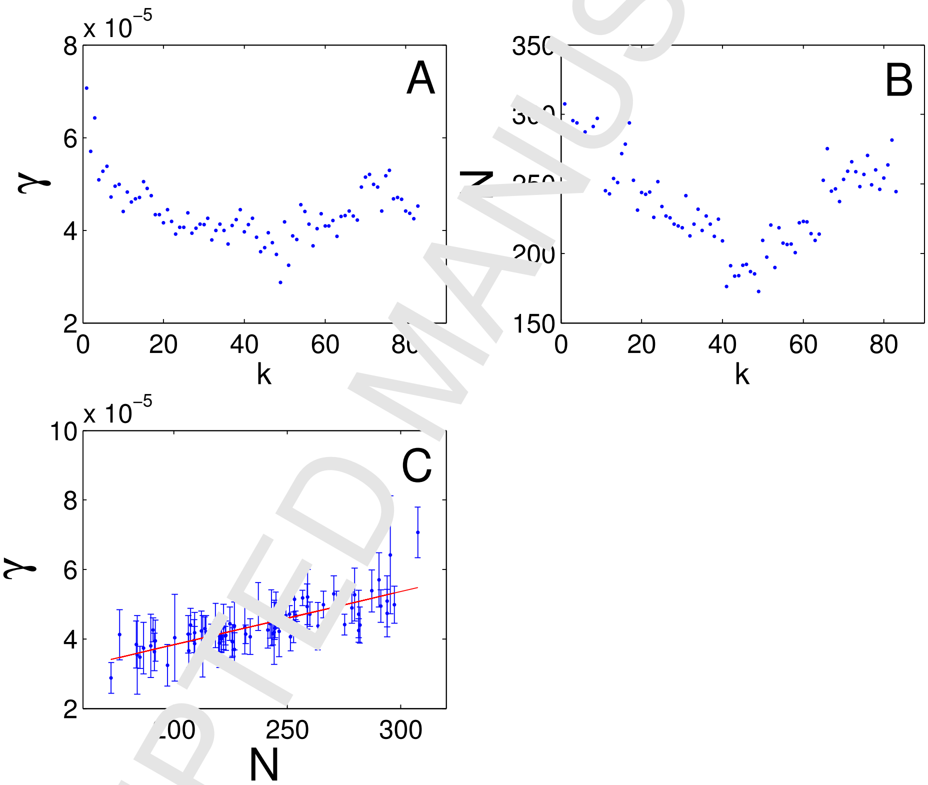

Figure 16: (Color online) (A) Volatility γ as a function of k for $\delta t = 300$ s. (B) Activity N as a function of k for $\delta t = 300$ s. (C) Scatter plot of volatility γ as a function of number of trades N. The points are averaged over the investigated period. All the plots are for simulated data with $w = 10$ s.