Centre for Computational Finance and Economic Agents

Working Paper Series

WP021-08

Rafael Velasco-Fuentes and Wing Lon Ng

Nonlinearities in Stochastic Clocks: Trades and Volume as Subordinators of Electronic Markets

June 2008

University of Essex

CCFEA

www.ccfea.net

Nonlinearities in Stochastic Clocks: Trades and Volume as Subordinators of Electronic Markets

RAFAEL VELASCO–FUENTES* and WING LON NG†

Centre for Computational Finance and Economic Agents (CCFEA)

University of Essex

Abstract

This paper shows that it is possible to recover normality of asset returns through a stochastic time change, where the appropriate operational time is determined through a function of the cumulative number of trades and/or the cumulative volume. Ané and Geman (2000) showed that the re-centered cumulative number of trades could be used as the appropriate stochastic clock of the market under which asset returns are virtually Gaussian. Using tick-data for FTSE 100 futures, we show that normality is not always recovered by conditioning on the re-centered number of trades, and instead demonstrate that a non-linear function of the number of trades and/or volume will provide a better stochastic clock of the market.

1 Introduction

Advances in *high-frequency finance* have been greatly encouraged by the recent surge in availability of large databases that contain information on every trade performed for a given asset. These data, usually referred to as tick-by-tick data (or simply tick-data), include vast amount of information not only about prices but also about timestamps and size of trade, which can prove valuable when studying the relationship between prices (or most commonly returns) and the activity in the market as measured by the number of transactions or the traded volume.

The positive relation between market activity and the volatility of asset returns has been widely documented over several decades: Early studies focused on the link between volume and volatility, e.g. Clark (1973) used traded volume in order to explain the non-normality of returns for cotton futures, while other models related trading volume and price movements to the arrival of new information (such as Harris (1982) and Tauchen and Pitts (1983)). Although later work also investigated the connection between asset prices and volume (see for example Gallant, Rossi, and Tauchen (1992), Schwert (1989), or – for a comprehensive review – see Karpoff (1987)), more recent research carried out by Jones, Kaul, and Lipson (1994) showed that the positive volume-volatility

*Financial support has been provided by the National Council for Science and Technology of Mexico (CONACYT) under grant 179358/194113. E-mail: *ravela@essex.ac.uk*

†E-mail: *wlng@essex.ac.uk*

relation is in fact due to the positive relation between between volatility and the number of transactions.

In the first study mentioned above, Clark (1973) put forward the idea that calendar time was not the natural time scale over which prices evolve. Although the notion of *operational* time[1] first emerged in the context of business cycles of macroeconomic variables (Burns and Mitchell, 1946) and in the investigation of 1920s hyperinflation (see Allais (1966) and Barro (1970)), it was Clark who first realized that price series evolve at "different rates during identical intervals of time" (Clark, 1973, p. 137). More recently Geman, Madan, and Yor (2000) proved that through the use of subordinated stochastic processes, asset prices can be thought of as continuous processes in an "economic" measure of time. The relation between volatility and number of transactions described above, was used by Geman and Ané (1996) and Ané and Geman (2000) in order to estimate the subordinated process under which asset prices are Brownian motion.

The main aim of this paper is to utilize the relationship between volatility and market activity in order to identify the appropriate process under which asset prices are Brownian motion. We allow for the volatility-activity relationship to be non-linear and estimate several parametric functional forms for this relation.

Using a set of 824 days of intraday tick-data on FTSE-100 index futures separated into four sub-periods, 10 minute intraday returns are computed in two different ways: using simple linear interpolation techniques, and by means of inhomogeneous time series operators as developed by Zumbach and Müller (2001). The relationship between the number of trades (traded volume) and the stochastic clock of the market is then estimated by finding the process for the trades (volume) under which the conditional returns are Gaussian.

The structure of the paper is as follows: Section 2 describes the general framework for constructing an homogeneous price series from "raw" tick-data, including a description of simple interpolation techniques and of convolution operators as presented in Dacorogna, Gençay, Müller, Olsen, and Pictet (2001). Section 3 contains a brief introduction to the market's operational time, including stochastic time changes and asset prices as Brownian motion. Details of the data are given in Section 4, followed by a discussion of intraday seasonalities in Section 5. Results and statistics of the conditional returns are presented in Section 6. Finally, a summary and concluding remarks are given in Section 7.

2 Operators and sampling schemes

In the following sections we introduce the mathematical framework that will be used to analyze the data, in particular the formalism of time-series operators and their application on irregularly spaced data.

A broad classification of time series can be done according to the spacing of their data points in time. Time series that are regularly spaced in time are known as *homogeneous*, while irregularly spaced series are termed *inhomogeneous*. The information recorded in tick-by-tick data arrives at random times and their time series are thus inhomogeneous. However, since most time series

[1] Referred to as economic time in the earliest studies, the recent literature in finance makes use of terms such as *business-*, *trading-*, *transaction-* or *operational*-time. In order to avoid confusion, only the terms economic and operational time will be used in this paper.

analysis methods are applicable only to homogeneous series[2], the raw tick-by-tick data has to be transformed into an homogeneous time series. Section 2.1 presents the most common way of constructing homogeneous time series from an inhomogeneous series via interpolation methods.

An alternative option to working with homogeneous time series is to utilise time series operators that transform an inhomogeneous time series into another inhomogeneous series. In Section 2.2 a particular group of this type of operators is introduced.

2.1 Usual sampling scheme

Most homogeneous time series that are created from raw (inhomogeneous) tick-data are constructed by sampling at a desired frequency and then using some *interpolation* method. If the (irregularly spaced) timestamps of the tick-by-tick data are given by times t_j with associated values[3] $z_j = z(t_j)$ then, by interpolation, we can create an homogeneous time series at times $t_0 + i\Delta t$ fixed to t_0, where the index i refers to the resulting homogeneous time series. Each sampling time $t_i = t_0 + i\Delta t$ will be bounded by two times t_j of the original inhomogeneous series, given by

$$\begin{aligned} t_{j'} \quad \leq \quad t_0 + i\Delta t \quad < \quad t_{j'+1} \quad , \qquad i = 1, 2, \ldots \quad , \\ \text{with} \quad j\,' = \max(j \,|\, t_j \leq t_0 + i\Delta t) \quad . \end{aligned}$$

Interpolation is then carried out between the points $(t_{j'}, z_{j'})$ and $(t_{j'+1}, z_{j'+1})$ following a specific interpolation technique as shown in Figure 1. The two most common interpolation methods are *previous-tick* interpolation (Wasserfallen and Zimmerman, 1985)

$$z_i \equiv z(t_i) = z_{j'} \tag{1}$$

and *linear* interpolation

$$z_i \equiv z(t_i) = z_{j'} + \frac{t_0 + i\Delta t - t_{j'}}{t_{j'+1} - t_{j'}} \left(z_{j'+1} - z_{j'} \right) \quad . \tag{2}$$

Each interpolation method has its advantages and disadvantages. While linear interpolation does not satisfy causality (i.e. at a time t_i it uses information from time $t_{j'+1}$, which lies ahead of t_i), previous-tick interpolation may produce spurious jumps in z when there are long periods of missing or no data. When the sampling interval Δt is large compared with the distance between ticks, both interpolation schemes will produce very similar homogeneous time series (Dacorogna et al., 2001). Therefore, throughout this paper we will only consider the case of linear interpolation. If the original (inhomogeneous) series $z(t_j)$ is the log-price of the asset, then the returns can be easily found as:

$$r(t_i) = z(t_i) - z(t_{i-1}) \quad . \tag{3}$$

Besides the time series of the log-prices z_j, tick-data usually contains information on the size of each trade (its volume), which we denote as $v_j \equiv v(t_j)$. The

[2] A well known exception is the ACD model introduced by Engle and Russell (1998).

[3] These associated values are usually log-prices, but they could refer to volume or other quantities.

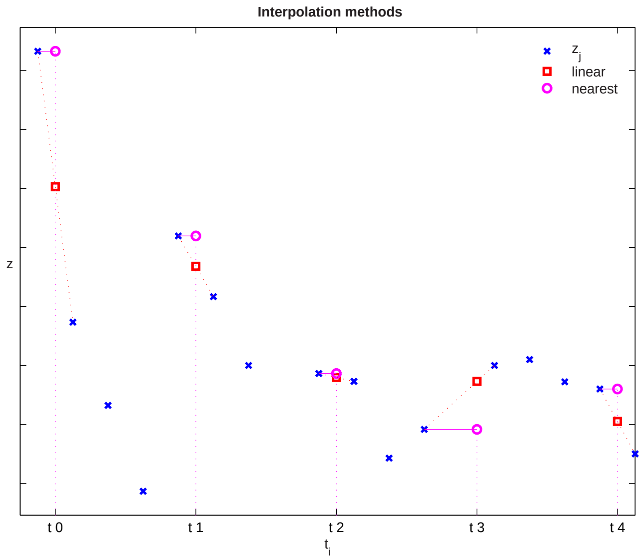


Figure 1: Interpolation methods. An inhomogeneous time series (blue) is sampled at regular time intervals in order to construct a regularly spaced (homogeneous) time series at times t_i. using previous-tick interpolation (magenta circles) and linear interpolation (red squares) schemes.

number of trades that have "accumulated" up to time t_j is simply $n'(t_j) = j$, while finding $v'(t_j)$ – the accumulated number of contracts traded up to time t_j – is straightforward with $v(t_j)' = \sum_{k=1}^{j} v(t_k)$.

Using any of the two interpolation schemes[4] presented above, $n'(t_i)$ – the values of an homogeneous time series for the number of trades up to times $t_0 + i\Delta t$ – can be found. The homogeneous series $v'(t_i)$ can be constructed in the same way.

The number of trades that take place during the interval $[t_{i-1}, t_i]$ is what we will refer to as the *cumulative* number of trades $N'(t_i)$, and is expressed as:

$$N'(t_i) = n'(t_i) - n'(t_{i-1}) \quad . \tag{4}$$

The cumulative volume $V'(t_i)$ can be found in a similar way:

$$V'(t_i) = v'(t_i) - v'(t_{i-1}) \quad . \tag{5}$$

Notice that equations (4) and (5) have a similar form to (3), the equation that is commonly used to obtain (log-)returns. This fact that will be of use in the next section when we make use of convolution operators.

2.2 Convolution operators

This section presents an alternative procedure to the one described in section 2.1, which can be used to construct homogeneous time series from inhomogeneous

[4]As mentioned earlier, in this paper only linear interpolation will be used.

tick-data. This methodology is based on the inhomogeneous time operators originally introduced by Zumbach and Müller (2001).

When dealing with tick-by-tick data, instead of focusing on pointwise values it makes more sense to deal with average values inside intervals and thus "the usual notion of return has to be changed" (Dacorogna et al., 2001, p. 51). For a homogeneous time series (e.g. data observed at intervals of size τ), a return is usually computed as the pointwise difference between the (log-) price observed at a given time t and the (log-) price of the previous observation (at time $t - \tau$). When working with inhomogeneous data (as tick-by-tick data) a better definition of the return could be the difference between the "average" price around time t and the "average" price around time $t-\tau$. With this idea in mind Zumbach and Müller (2001) developed a framework of efficient operators, where an operator Ω that acts on a generic (homogeneous or inhomogeneous) time series z is denoted $\Omega[z]$. Their work focused on operators that have the following properties:

- Linearity. Where operators satisfy: $\Omega[z_1 + c\, z_2] = \Omega[z_1] + c\, \Omega[z_2]$
- Time–translation invariance. With: $\Omega[z(t-\Delta)](t) = \Omega[z(t)](t-\Delta t)$, where $\Omega[z](t)$ is the value of $\Omega[z]$ at time t.
- Causality. Where $\Omega[z](t)$, i.e. the value of the operator at time t, depends only on information up to time t.

An operator with the above properties can be represented by the convolution with a kernel $\omega(t)$:

$$\Omega[z](t) \;=\; \int_{-\infty}^{t} dt'\, \omega(t-t')\, z(t') \quad . \tag{6}$$

The simplest of the linear operators proposed by Zumbach and Müller (2001) is the exponential moving average (EMA), which has an exponentially decaying kernel[5] given by:

$$\mathrm{ema}[\tau](t) = \frac{e^{-t/\tau}}{\tau} \quad , \tag{7}$$

where τ is the characteristic time range (i.e. the time interval). In practice, there is no need to compute the convolution integral in (6), instead one can use a simple iterative formula first proposed by Müller (1991) for the EMA:

$$\mathrm{EMA}[\tau; z](t_j) = \mu\, \mathrm{EMA}[\tau; z](t_{j-1}) + (\nu - \mu)\, z_{j-1} + (1-\nu)\, z_j \tag{8}$$

with

$$\alpha = \frac{t_j - t_{j-1}}{\tau} \quad \text{and} \quad \mu = e^{-\alpha} \quad .$$

Since the definition of the convolution in (6) makes use of an integral, $z(t')$ should be a continuous function of time. However, since the time series of interest are not continuous, an interpolation between points must be used in

[5]To distinguish between the operators and their respective kernels, the former are always capitalised.

order to compute the convolution. The type of interpolation method that is used in equation (8) is defined by the parameter ν:

$$\nu = \begin{cases} 1 & \text{previous-point interpolation} \\ (1-\mu)/\alpha & \text{linear interpolation} \\ \mu & \text{next-point interpolation} \end{cases}$$

Due to the EMA's computational efficiency, it serves as the basis of other more complex operators. Iterations of the basic EMA operator can provide us with a set of iterated exponential moving average operators $\text{EMA}[\tau, n]$, with kernel:

$$\text{ema}[\tau, n](t) = \frac{1}{(n-1)!} \left(\frac{t}{\tau}\right)^{n-1} \frac{e^{-t/\tau}}{\tau} \quad ,$$

where n denotes the number of iterations used. Once more the iterated EMA operator does not need to be calculated through a convolution, it can instead be computed using the simple recursive formula:

$$\text{EMA}[\tau, n; z] = \text{EMA}[\tau; \text{EMA}[\tau, n-1; z]] \quad , \tag{9}$$

with $\text{EMA}[\tau, 1; z] = \text{EMA}[\tau; z]$.

A last point to mention is that since the kernels have an exponential tail for large t, when applying an **EMA** operator to a time series, a build–up time will required and a lag with respect to the original time series will be introduced (see top panel in Figure 2).

Through the use of the iterated **EMA** operator of equation (9), a low-noise differential operator that measures the difference between an average value now and an average value in the past can be used in order to obtain the "returns" of our inhomogeneous tick-data series. Zumbach and Müller (2001) proposed the differential operator $\mathbf{\Delta}[\tau]$, as a suitable operator to measure returns:

$$\mathbf{\Delta}[\tau] = \kappa\left(\text{EMA}[a\tau, 1] + \text{EMA}[a\tau, 2] - 2\ \text{EMA}[ab\,\tau, 4]\right) \quad , \tag{10}$$

with $\kappa = 1.22208$, $b = 0.65$ and $a^{-1} = \kappa(8b - 3)$. Where the value of κ ensures that the integral of the kernel from the origin to the first zero is equal to one; the value of b is fixed in order to get a short tail and a is fixed by the normalisation condition[6].

By applying the differential operator in (10) to the tick-data price series, a new inhomogeneous time series of "economic time" (log-) returns can be obtained. This new series can then be sampled at a given frequency in order to construct an homogenous time series of returns.

A comparison of the returns constructed using usual interpolations schemes described in Section 2.1 and the returns obtained by sampling on the homogeneous series obtained via the differential operator is shown in the bottom panel of Figure 2. The inhomogeneous time series of returns via the differential operator is shown as a light–blue line, by sampling this time series every 10 minutes, an homogeneous time series of returns is obtained (dark blue dots). The series of 10min returns computed using the usual method (with linear interpolation) is shown as red circles.

[6]Chosen so that for a constant function $c = c(t)$, $\mathbf{\Delta}[\tau; c] = 0$ and $\mathbf{\Delta}[\tau; t] = \tau$.

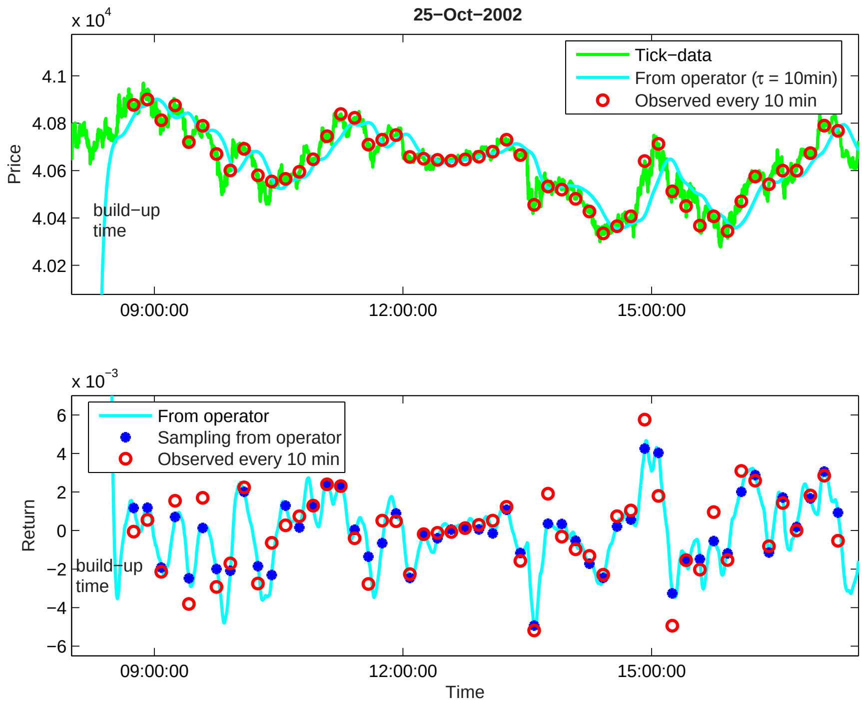


Figure 2: Prices and returns series for 25 October 2002. Top: Tick-data prices (green line); prices after using the EMA operator (light-blue line) and homogeneous price series (red circles) constructed by linear interpolation by sampling every 10 minutes. Bottom: Returns constructed via the differential operator with $\tau = 10$min (light-blue line). Homogeneous time series of returns by sampling the differential operator at 10 minute intervals (blue dots), and 10 minute series of returns constructed following the usual sampling scheme (red cirles). The "build-up" time is clearly visible.

Notice that although the two homogeneous series of returns (red circles and blue dots) move together, they are not always the same. The returns sampled from the differential operator incorporate more information from the original raw tick–data and thus reflect variations in the price that the other series of returns does not capture.

Finally, by applying the differential operator in (10) to the series $n'(t_j)$ we can obtain an inhomogeneous time series $N'_{\text{op}}(t_j)$, which can then be sampled at times t_i in order to obtain an homogeneous time series for the cumulative number of trades $N'_{\text{op}}(t_i)$. Once more, by sampling from the time series after the operator has been applied, more information from the original (tick-data) series will be incorporated than if the usual interpolation method had been used. In order to obtain $V'_{\text{op}}(t_i)$ – an homogeneous time series for the cumulative volume – the same procedure can be applied to $v'(t_j)$. The series $V'_{\text{op}}(t_i)$ will once more incorporate more information from the original raw series than $V'(t_i)$ (the cumulative volume using the usual linear interpolation method described in Section 2.1).

The top panel in Figure 3 shows the inhomogeneous time series for the cu-

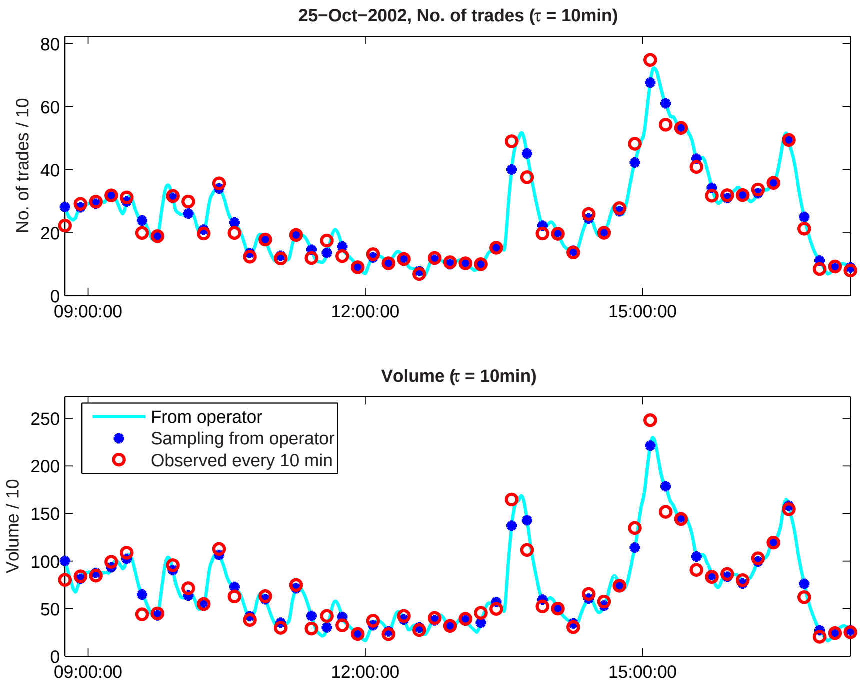


Figure 3: Number of trades and volume for 25 October 2002. Top: (Cumulative) Number of trades with τ = 10min calculated using the differential operator on the (accumulated) number of trades (light blue), by sampling of the operator (blue dots) and by observing the number of trades in every 10 minute interval (red circles). Bottom: (Cumulative) Volume with τ = 10min. Calculated by applying the differential operator on the (accumulated) volume (light-blue line); by 10 minute sampling of the operator (blue dots) and by counting the number of contracts trades in every 10 minute interval (red circles).

mulative number of trades $N'_{\mathrm{op}}(t_j)$ (light–blue line); $N'(t_i)$, the homogeneous time series constructed by a simple sampling technique (red circles) and $N'_{\mathrm{op}}(t_i)$, the homogeneous series obtained by sampling of the operator (blue dots). The inhomogeneous time series for volume $V'(t_j)$ (light–blue line) and the homogeneous $V'(t_i)$ and $V'_{\mathrm{op}}(t_i)$ are shown in the bottom panel (red circles and blue dots respectively). Both panels shown are constructed from intraday tick–data for the 25 October 2002, with the homogeneous time series starting at 8:45 and finishing at 17:15.

3 Operational time of the market

In this section the mathematical framework of stochastic time changes is introduced, along with the definition of what constitutes the appropriate operational time of the market in relation with the number of trades and volume.

3.1 Asset price under a stochastic time change

Let $x^{(s)}$ be a discrete stochastic process indexed by a discrete variable. Such a process may be expressed as:

$$x^{(s)} = \{x(0), x(1), \ldots, x(s), x(s+1), \ldots\} \quad , \tag{11}$$

where $x(s)$ is the observed log-price of the asset in operational time, e.g. the price after s units of operational time.

Consider now a discrete process $x^{(t)}$ made of observations of the price x at (calendar) times $t_1, t_2, t_3, \ldots, t, \ldots$, with $t_1 < t_2 < t_3 < \cdots < t < \ldots$, i.e. a process given by

$$x^{(t)} = \{x(t_1), x(t_2), x(t_3), \ldots, x(t), \ldots\} \quad , \tag{12}$$

and define a positive and strictly increasing stochastic process

$$\{\vartheta(t_1), \vartheta(t_2), \vartheta(t_3), \ldots, \vartheta(t), \ldots\} \quad , \tag{13}$$

where $\vartheta(t_i)$ measures the units of economic time that have passed up to time t_i.

In general, up to a time t, s units of economic time will have passed, i.e.

$$s = \vartheta(t) \quad , \tag{14}$$

where (14) denotes a random variable that defines the stochastic process in (13). The processes described by (11) and (12) are related by

$$x^{(t)}(t) = x^{(s)}(\vartheta(t)) \quad . \tag{15}$$

At time t, $\vartheta(t)$ units of economic time will have passed, so at this time processes $x^{(t)}$ and $x^{(s)}$ will have the same value. It is straightforward to observe that intraday calendar returns will also be related to operational time via

$$r(t_i) = \Delta x^{(s)}(\vartheta(t_i)) \quad , \tag{16}$$

where $\Delta x^{(s)}(\vartheta(t_i)) \equiv x^{(s)}(\vartheta(t_i)) - x^{(s)}(\vartheta(t_{i-1}))$.

In the terminology of subordinated stochastic processes, the non-stationary process $x^{(t)}$ (the asset price observed at equidistant time intervals) is known as the *latent process*, and is said to be *subordinated* to the *parent process* $x^{(s)}$ (the asset price under operational time), and *directed* by the *subordinator*[7] $s = \vartheta(t)$ (Feller, 1971).

3.2 An appropriate definition of operational time

The original subordinated processes used by Clark (1973) represented asset prices and time by two independent geometric Brownian motions respectively. His findings have been generalised by Geman et al. (2000), who used pure jump processes of finite variation to model asset returns. These processes can be expressed as Brownian motion evaluated at random times with the added generality that the process of the stochastic clock need not be independent of the price process.

[7] The subordinator is sometimes also referred to as the *directing process*.

It is well known that the existence of a risk-neutral probability measure under which discounted asset prices are martingales is a result of the no-arbitrage assumption, and that under the real-world probability asset prices must then be semimartingales. Monroe (1978) stated an essential theorem in which he showed that any semimartingale is equivalent to a time change of Brownian motion. If the asset price is a semimartingale, then the asset log-price process in calendar time, $x^{(t)}(t)$, will also be a semimartingale, and following Monroe's theorem it may be written as:

$$x^t(t) \equiv W(\vartheta(t)) \quad , \tag{17}$$

where $W(\cdot)$ is a Wiener process.[8]

Geman et al. (2000) compared (17) to (15) and proposed that the appropriate choice for operational time will be the one under which the asset prices are Brownian motion. It follows from this statement that in order to find the appropriate stochastic clock for operational time one must identify the process under which the distribution of returns is Gaussian.

With a vast amount of literature focusing on the relation between asset prices and traded volume, a natural choice to describe the "activity" of financial markets would be to use the volume traded over an interval of time. Nevertheless, studies by Jones et al. (1994) singled out the number of transactions and not their size as the main factor that determines the volatility of the market (as mentioned in Section 1). Work done by Ané and Geman (2000) showed that the appropriate subordinator of operational time for asset returns could be approximated by a linear function of the cumulative number of trades, with asset returns conditional on the re-centered (cumulative) number of trades being virtually Gaussian. However, other studies (e.g. Murphy and Izzeldin (2006) and Velasco–Fuentes and Chourdakis (2007)) have shown that conditioning on the number of trades does not always recover normality.

In the rest of this paper no assumption will be made of whether it is the (cumulative) number of trades *or* the (cumulative) volume that best relate to the operational time of the market. Instead, the two alternatives will be investigated and a third option, where *both* trades and volume are considered simultaneously, will also be examined.

3.3 Number of trades and volume as subodinators

The distribution of returns conditional on the (re-centered) cumulative number of trades can be computed using the following re-normalisation

$$\breve{r}_N(t_i) = \frac{r(t_i)}{\sqrt{\breve{\beta}_N \, N'(t_i) + \breve{\gamma}_N}} \quad , \tag{18}$$

where $\breve{\beta}_N$ and $\breve{\gamma}_N$ are constants estimated such that the conditional returns $\breve{r}_N(t_i)$ are approximately Gaussian with variance one. Adding the constant to the number of trades re-centers their distribution, while the $\breve{\beta}_N$ is introduced in order for the conditional returns to have unit variance.

As mentioned in Section 3.2 instead of the number of trades, an alternative is to use $V'(t_i)$, i.e. the volume traded during the interval $[t_{i-1}\,, t_i]$. The

[8] Brownian motion with zero drift and variance one.

returns conditional on the (re-centered) volume can then be found using the normalisation:

$$\breve{r}_V(t_i) = \frac{r(t_i)}{\sqrt{\breve{\beta}_V\, V(t_i) + \breve{\gamma}_V}} \quad , \tag{19}$$

where the constants $\breve{\beta}_V$ and $\breve{\gamma}_V$ are once more estimated such that the conditional returns $\breve{r}_V(t_i)$ are approximately Gaussian with variance one.

In a more general form, the re–normalisations in (18) and (19) can be expressed as the calendar returns divided over the square root of a function $f(\cdot)$,

$$\breve{r}(t_i) = \frac{r(t_i)}{\sqrt{f(\cdot)}} \quad , \tag{20}$$

where $f(\cdot)$ can be either a function of N or V (the number of trades or the volume), or of both. The simplest case being when $f(\cdot)$ is a linear function of N or V as in equations (18) and (19).

As stated in Section 3.2, conditioning on the re-centered cumulative number of trades does not always recover normality (see Table 3). In order to recover normality and find the appropriate process for operational time, Velasco–Fuentes and Chourdakis (2007) proposed a non-linear relation between the number of trades and the stochastic clock of the market, and estimated a semi-parametric function of the number of trades under which conditional returns were much closer to Gaussian than those conditional on the re-centered number of trades. To determine the appropriate subordinator for the stochastic clock of the market, several parametric forms for $f(\cdot)$ will be estimated. The conditional returns will be then found via (20) and tested for normality (see section 6).

4 Data

Our database consists of tick-data for FTSE 100 index futures[9] spanning the period from 2 January 2001 to 17 June 2004, which comprises a total of 824 trading days. This interval includes a wide-range of market conditions and activity, from days of high market commotion and volatility to days of very low activity. From a daily point of view, the maximum number of trades recorded during a particular day is 31,440 (on 18 March 2003), with a minimum of 2,226 (on 27 December 2001). The average number of trades per day for the whole sample is 12,185.99, with a standard deviation of 4,666.63 trades per day. In the case of the volume, the maximum traded in a single day was 124,870 contracts (on 24 July 2002), the minimum traded was 10,410 (on the 18 February 2002). The volume per day had an average of 41,342.71 and a standard deviation of 17,346.05 contracts per day.

Using the differential operator of equation (10) with $\tau = 10$min, the inhomogeneous series of intraday returns, $r(t_j)$, was calculated for each trading day. In order to avoid the build-up time of the operator all points with timestamps prior to 8:45:00 of the inhomogeneous series of returns are dropped, with the last 15 minutes of the trading day also removed in order to avoid closing effects. The resulting inhomogeneous series are then sampled at 10 minute intervals

[9]FTSE 100 futures are traded between 8:00 and 17:30.

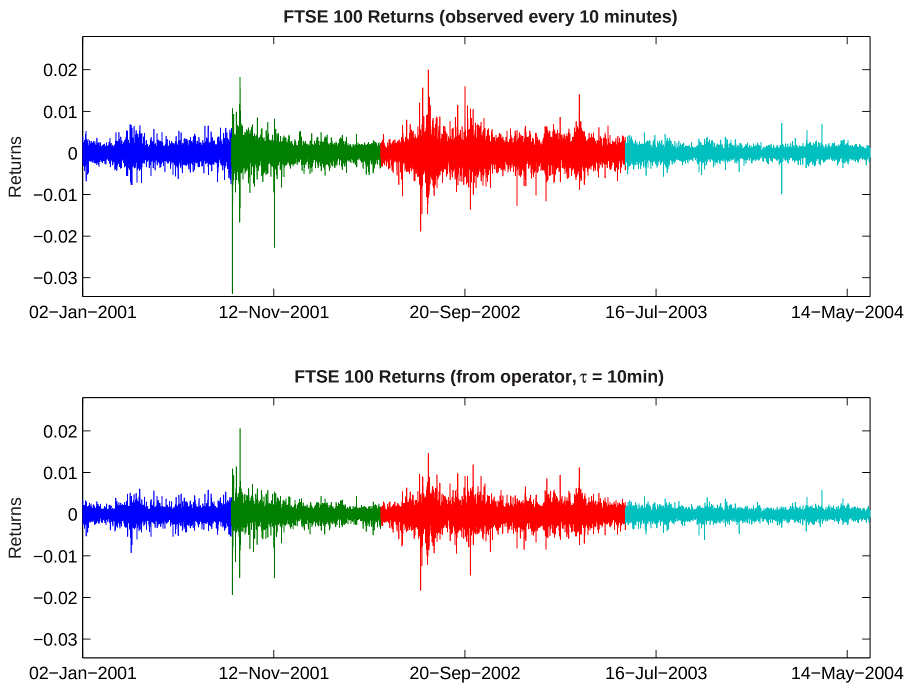


Figure 4: FTSE 100 Futures Index 10 minute returns. The four sub-periods are clearly shown: An initial period of 156 days before 11 September 2001 (blue), a second period that includes 11 September 2001 (green), followed by a period of 256 days of high-volatility (red) and finally 256 less volatile days (light blue). Bottom: Returns obtained through the usual method of sampling prices (at 10 minute intervals). Bottom: Returns constructed using sampling on the differential operator (with $\tau = 10$min).

in order to create the homogeneous time series of returns $r_{\text{op}}(t_i)$. Using the usual sampling method (at intervals of 10 minutes) with linear interpolation described in Section 2.1 the intraday homogeneous series of returns $r(t_i)$ are also computed.[10]

In a similar manner, two homogeneous series are constructed for the (cumulative) number of trades: $N'(t_i)$ by linear interpolation of observations every 10min, and $N'_{\text{op}}(t_i)$ by applying the differential operator as described in Section 2. Finally, the time series for the (cumulative) volume $V'(t_i)$ and $V'_{\text{op}}(t_i)$ are also constructed. The descriptive statistics of these homogeneous time series are shown in Table 1.

For the main analysis the data was divided into four sub-periods: The first one spanning all data prior to 10 September 2001 (consisting of 156 days). The second subperiod includes also 156 days of data starting from 11 September 2001 and concluding on 7 May 2002. The third and fourth subperiods include 256 days of data each, the former starting on 7 May 2002 and ending on 30 May 2003, while the last sub-period expands from 2 June 2003 to 17 June 2004.

The descriptive statistics of the 10 minute intraday returns for each subpe-

[10] At the same times t_i, i.e. at 10min intervals starting at 8:45 and finishing at 17:15

riod are shown in Table 2, and a graphical representation of the returns $r(t_i)$ and $r_{\mathrm{op}}(t_i)$ shown in figure 4. As expected the returns in the second subperiod (green) have the highest skewness and kurtosis since that period contains the events of 11 September 2001. The third subperiod (red) is the most volatile, while the last (light-blue) is the least volatile. The Jarque-Bera (JB) test statistics in Table 2 show that these returns are far from Gaussian[11] in all cases.[12]

Before trying to find the process under which the asset returns are Gaussian, one must first take into account any deterministic intraday patterns as explained in the next section.

5 Intraday seasonality and volatility

A well established characteristic of activity[13] in most markets is that it is higher during the hours immediately after opening and closing than in the middle of the trading day (see for example Brock and Kleidon (1992) and Bollerslev and Domowitz (1993)). This *intraday seasonality* effect[14] extends to other asset characteristics, most importantly to intraday volatility.

The intraday seasonality pattern will be estimated for the three quantities of interest: intraday volatility, cumulative number of trades and cumulative volume. The intraday diurnal component of the volatility is computed as the average of the squared returns at given times during the trading day measured over all days in a sample.

For $r(t_i)$, i.e. the returns obtained by using linear interpolation as in (3), the diurnal component of the volatility $\phi(t_i)$, is computed in every subperiod as the square root of the average of the squared returns at (equidistant) times t_i:

$$\phi(t_i) = \left(\mathrm{E}[r^2(t_i)]\right)^{1/2} \quad ,$$

where the expectation is taken over all days in the subperiod. $R(t_i)$, the "deseasonalised" returns, i.e. the returns from which the deterministic diurnal component has been removed are then calculated in a multiplicative way:

$$R(t_i) = \frac{r(t_i)}{\phi(t_i)} \quad . \tag{21}$$

Alternatively, $\phi_{\mathrm{op}}(t_j)$ – the deterministic diurnal component of the volatility at non-equidistant times t_j can be used. The inhomogeneous time series of intraday prices of each day are weighted and incorporated into a single inhomogeneous series that is then passed through the differential operator and its result squared in order to obtain $\phi_{\mathrm{op}}(t_j)$. The diurnal component of the volatility $\phi_{\mathrm{op}}(t_i)$ at equidistant times t_j can then be calculated by interpolation and the deseasonalised returns $R_{\mathrm{op}}(t_i)$ found as

$$R_{\mathrm{op}}(t_i) = \frac{r_{\mathrm{op}}(t_i)}{\phi_{\mathrm{op}}(t_i)} \quad . \tag{22}$$

[11]Throughout this paper we focus only on the skewness and kurtosis aspects of normality, without taking into account autocorrelation properties of returns.

[12]The critical value of the JB test is 9.21 (5.99) at 1% (5%) significance level

[13]This phenomenon holds both for the number of trades and for volume traded.

[14]Other effects, such as the seasonality related to the day of the week, will not be dealt with in this paper.

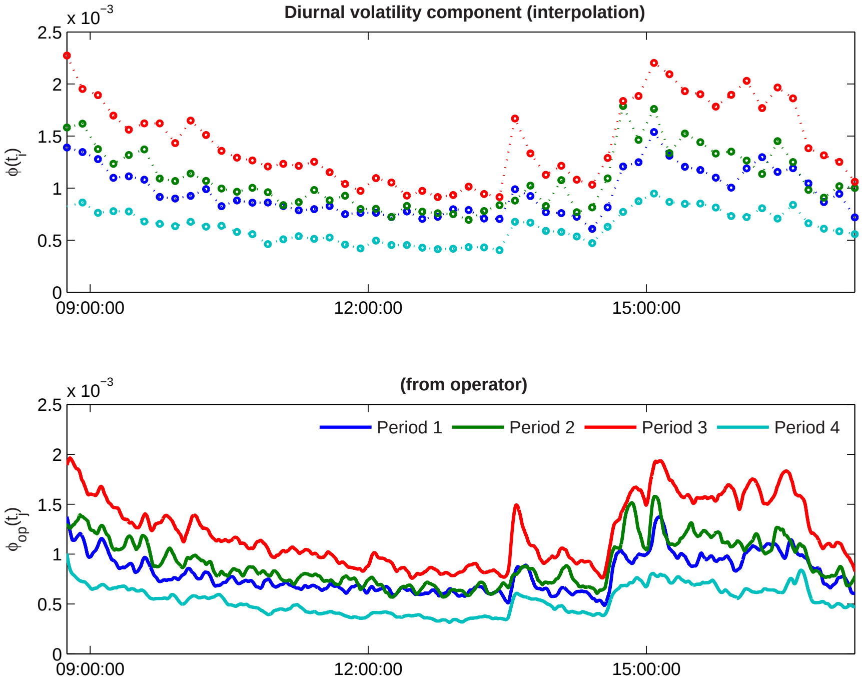


Figure 5: Diurnal component of the squared returns. Top: Diurnal components of the squared returns as computed by usual interpolation scheme with sampling at 10 minute intervals. The overall level of the diurnal component is clearly lowest during the fourth sub-period (light blue), with the highest being the one of the third subperiod (red). Bottom: The expected component of the squared returns as computed by the differential operator for the four different sub-intervals.

The diurnal components for the (cumulative) number of trades $\phi_N(t_i)$ and (cumulative) volume $\phi_V(t_i)$, can be found in a similar way as for the volatility and is then used to find their deseasonalised counterparts:

$$N(t_i) = \frac{N'(t_i)}{\phi_N(t_i)} \quad \text{and} \quad V(t_i) = \frac{V'(t_i)}{\phi_V(t_i)} \quad . \tag{23}$$

Using the differential operator the deseasonalised number of trades and volume will be given by

$$N_{\text{op}}(t_i) = \frac{N'_{\text{op}}(t_i)}{\phi_{N,\,\text{op}}(t_i)} \quad \text{and} \quad V_{\text{op}}(t_i) = \frac{V'_{\text{op}}(t_i)}{\phi_{V,\,\text{op}}(t_i)} \quad . \tag{24}$$

Figure 5 shows the deterministic diurnal pattern of the intraday volatility for the four different subperiods. The top panel shows $\phi(t_i)$, the volatility pattern computed using the homogeneous time series of returns, while the bottom panel shows $\phi_{\text{op}}(t_j)$, i.e. the deterministic volatility pattern computed using the differential operator. A "U–shaped" pattern is clearly noticeable for all subperiods, with the level of volatility being high during the early hours of the trading day and decreasing as lunch time approaches, followed by a spike at the time of the opening of the US markets and increased volatility thereafter until it diminishes again prior to closing.

The diurnal component of the number of trades (Figure 6) and of the volume (Figure 7) show a pattern similar to the one of the intraday volatility, i.e. decreasing activity towards lunch time with a sudden increase at the opening of the US market and a period of high activity in the hours just before closing.

The descriptive statistics of the deseasonalised returns $R(t_i)$ and $R_{\mathrm{op}}(t_i)$ are shown in Table 2. Taking into consideration the diurnal component of volatility considerably decreases the skewness and kurtosis of the returns in almost every sub–period. However, these returns are still far from being Gaussian as can be seen from the excessively high value of the JB statistic. The conditional returns as mentioned in the next section, will be computed using these deseasonalised returns, number of trades and volume.

6 Conditional returns

The returns conditional on the (re-centered) number of trades can be obtained via equation (18), where the deseasonalised returns, $R(t_i)$, and the number of trades, $N(t_i)$, are used in the minimisation problem:

$$\min_{\breve{\beta},\breve{\gamma}} \mathcal{JB}(\breve{\beta},\breve{\gamma};R_N(t_i)) \quad \text{where} \quad R_N(t_i)=\frac{R(t_i)}{\sqrt{\breve{\beta}_N N(t_i)+\breve{\gamma}_N}} \tag{25}$$

subject to $\mathrm{Var}[R_N(t_i)] = 1$, where $\mathcal{JB}(R_N(t_i))$ gives the JB test statistic of the returns $R_N(t_i)$. The returns conditional on volume can also be obtained in a similar manner by replacing the number of trades with the (deseasonalised) volume $V(t_i)$ in equation (25). The parameters $\breve{\beta}_N$ and $\breve{\gamma}_N$ in (25) represent a line that describes the relationship between the number of trades and the process under which the returns are closest to Brownian motion, i.e. the process that determines the stochastic clock of the market.

Table 3 shows the skewness, kurtosis and values of the JB statistic of the conditional returns R_N and R_V found by minimising the JB statistic as in (25). These conditional returns have lower skewness and excess kurtosis than their calendar counterparts (compare with Table 1). Nevertheless, although conditioning on the re-centered number of trades and volume yields returns closer to Gaussian, normality is only recovered in a few cases.[15]

6.1 Non-linear relationship

As discussed in Section 3.3, a more general approach can be achieved by letting the function of the number of trades (or volume) take on forms other than linear. The most general approach is best described through the minimisation

$$\min_{\hat{\Theta}} \mathcal{JB}(\hat{\Theta};R_{N,V}) \quad \text{where} \quad R_{N,V}=\frac{R}{\sqrt{f(N,V;\hat{\Theta})}} \quad , \tag{26}$$

[15]The hypothesis of normality can not be rejected when the JB statistic is lower than 9.21 (at 1% significance level). These cases are shown in bold in Table 1.

subject to $\mathrm{Var}[R_{N,V}] = 1$. Where R, N and V are the deseasonalised calendar returns, number of trades and volume[16] respectively; and $\hat{\Theta}$ is a vector of parameters of the function of the number of trades and volume $f(N, V)$.

Besides the particular cases of the linear functions:

- $f_{\text{lin}}(N) = \breve{\beta}_N N + \breve{\gamma}_N$, and
- $f_{\text{lin}}(V) = \breve{\beta}_V V + \breve{\gamma}_V$

mentioned above, we propose separate quadratic forms for N and V:

- $f_{\text{quad}}(N) = \hat{\alpha}_N N^2 + \hat{\beta}_N N + \gamma_N$, and
- $f_{\text{quad}}(V) = \hat{\alpha}_V V^2 + \hat{\beta}_V V + \gamma_V$.

With their results presented in Table 3. The conditional returns constructed in this way will be, by construction, at least as close to Gaussian as those conditioned on the linear function, since the latter is a quadratic with $\hat{\alpha} = 0$. As discernible in Table 3, these conditional returns are approximately Gaussian for the two sub-periods with less volatility.

If the relation between the stochastic clock of the market and the number of trades (or volume) is linear, whenever there is a change in trading activity[17] the "clock" of the market will only be affected by the magnitude of this change, e.g. an increment in activity from 10 trades per minute to 11 trades per minute will have the same effect on the "clock" of the market as an increase from 110 to 111 trades per minute due to the constant slope of the linefootnoteThe same is true if the volume is used instead of the trades. By allowing for the relation between the number of trades (or the volume) and the stochastic clock of the market to be non-linear, the relative size of the change can be taken into account. Going back to the previous example, an increment in activity from 10 to 11 trades/min can now have a different effect on the clock of the market than the increment from 110 to 111 trades/min, i.e. an increase in activity has a greater effect on the operational time when activity is already high than when it is low. The estimated functions for the linear and quadratic forms mentioned above are shown (together with their estimated parameters) in Tables 5 to 8.

In order to allow for asymmetric responses of the market to upward and downward price movements, i.e. for the possibility that the market's clock will react differently to changes in activity when returns are positive than when they are negative, the function of the number of trades can be constructed having two separate components:

$$f_{\text{lin}}^{\mp}(N) = \begin{cases} \breve{\beta}_N^- N(t_i) + \breve{\gamma}_N^- & \text{if} \quad R(t_i) < 0 \\ \breve{\beta}_N^+ N(t_i) + \breve{\gamma}_N^+ & \text{if} \quad R(t_i) \geq 0 \end{cases} , \tag{27}$$

where $\breve{\beta}_N^{\mp}$ and $\breve{\gamma}_N^{\mp}$ are parameters that must be estimated by minimising the JB statistic as described in (26). A similar function can also be written for the volume as $f_{\text{lin}}^{\mp}(V) = (\breve{\beta}_V^- V + \breve{\gamma}_V^-)\,\mathbf{1}_{\{R(t_i)<0\}} + (\breve{\beta}_V^+ V + \breve{\gamma}_V^+)\,\mathbf{1}_{\{R(t_i)\geq 0\}}$. Returns conditioned on these functions will have JB statistics at least as low as the

[16]The same minimisation problem can be solved for $R_{\text{op}}, N_{\text{op}}$ and V_{op}, yielding conditional returns $R_{N,V}^{\text{op}}$

[17]Measured as an increase (or decrease) of the number of trades (or volume) per time interval.

simple linear functions $f_{\text{lin}}(\,\cdot\,)$, since the latter is a specific case of (27) when $\breve{\beta}^- = \breve{\beta}^+$ and $\breve{\gamma}^- = \breve{\gamma}^+$. The skewness, kurtosis and JB statistics of the returns conditional on $f^{\mp}_{\text{lin}}(\,\cdot\,)$ can be seen in Table 4, with the values of the estimated parameters and plots of the functions shown in Tables 5 to 8. As expected, these returns are closer to Gaussian than those conditioned on the simple linear function. One can also observe that dealing with negative and positive returns separately has substantially lowered the skewness of the conditional returns in most sub-periods when compared with those from the simpler linear function.

The quadratic function can also be modified so that the effect of the changes in activity on the clock of the market will differ depending on the sign of the returns, this will be addressed by using the function

$$f^{\mp}_{\text{quad}}(N) = \begin{cases} \hat{\alpha}^-_N N(t_i)^2 + \hat{\beta}^-_N N(t_i) + \hat{\gamma}^-_N & \text{if} \quad R(t_i) < 0 \\ \hat{\alpha}^+_N N(t_i)^2 + \hat{\beta}^+_N N(t_i) + \hat{\gamma}^+_N & \text{if} \quad R(t_i) \geq 0 \end{cases} \quad , \tag{28}$$

which by construction will give conditional returns that are at least as close to normal as those obtained when conditioning on the simple quadratic function $f_{\text{quad}}(\,\cdot\,)$. The skewness, kurtosis and JB statistics of the returns conditioned on this type of function are presented in Table 4 and the estimated functions (along with the estimated parameters) are shown in Tables 5 to 8.

6.2 Returns conditional on trades and volume

It was mentioned in Section 6.1 that the appropriate process under which the returns are Gaussian could be a function of both the number of trades N *and* of the volume V simultaneously (i.e. $f(N, V)$ in equation 26). However, so far we have only considered the cases when the relation is given by either a function of trades *or* by a function of volume. We can relax this assumption and consider the following function of both:

$$f(N, V) = \tilde{\alpha}_N N^2 + \tilde{\alpha}_V V^2 + \tilde{\alpha}_{NV} N \cdot V + \tilde{\beta}_N N + \tilde{\beta}_V V + \tilde{\gamma} \quad , \tag{29}$$

where the parameters are once more estimated according to (26). Yet again, returns conditional on (29) will be closer to Gaussian than those conditioned on $f_{\text{lin}}(\,\cdot\,)$ and $f_{\text{quad}}(\,\cdot\,)$, since the former two functions are special cases of $f(N, V)$. The skewness, kurtosis and JB statistics of the returns conditional on $f(N, V)$ are given on Table 4.

7 Summary and conclusions

The notion that financial markets have their own "operational" time different from the usual calendar time has been examined throughout this paper by focusing on the relationship between market activity – proxied by number or trades or volume – and the processes under which returns are normal. These processes are assumed to be determined by a function of the (cumulative) number of trades and/or the (cumulative) volume, with several parametric forms of this function where estimated.

Even though calendar returns are evidently non-normal, with high skewness and excess kurtosis, these same returns conditional on the (re-centered) number of trades or (re-centered) volume are closer to Gaussian. Other processes under the number of trades and/or volume where also investigated, including the case when a change in activity has different effects on the clock of the market depending on the sign of the return associated with that change, and the case when the effect is dependent not only on the size of the change but on its overall level.

As indicated by the results, skewness of the conditional returns can be greatly reduced by allowing the effect of a change in activity to be asymmetric, while kurtosis can be dealt with by making the stochastic clock of the market react more to changes in activity when activity is high than when it is low. A last approach was to utilize a function of both number of trades and volume in order to find the conditional returns.

Tick-data of FTSE 100 index futures was used in order to construct 10 minute returns spanning an interval of over three years. This data was separated into four sub-periods with very different levels of volatility, skewness and kurtosis. Conditioning on the number of trades and/or volume via different functions produces returns that are Gaussian only for the two least volatile sub-periods, however, normality was not recovered in the other two sub-periods. These last results could indicate that the link between the operational time of the market and the number of trades and volume during turbulent times could be more complex than the relations proposed in this paper. Nonetheless, the simple relationships that have been forward succeed in explaining possible connections between changes in market activity and skewness and kurtosis of returns.

Finally, a last remark must be made regarding the autocorrelation of returns. Although the fact that asset returns possess significant autocorrelation properties[18] was not addressed in this study, one could conjecture that this autocorrelation arises as a result of autocorrelation in the process followed by the number of trades (volume), i.e. intervals with high (low) market activity –as gauged by the number of trades (volume)– will be followed by other intervals of high (low) activity. This final observation leads us to believe that future research should not only improve the understanding of the number of trades and volume as factors that determine the operational time of the market, but also towards developing appropriate models for the process of number of trades (and volume) that will also explain other autocorrelation properties.

[18] Of particular importance is the autocorrelation of the squared returns (or of their absolute values).

References

Allais, M. (1966). A restatement of the quantity theory of money. *A.E.R. 56*, 1123–1157.

Ané, T. and H. Geman (2000). Order flow, transaction clock, and normality of asset returns. *The Journal of Finance 55*, 2259–2284.

Barro, R. J. (1970). Inflation, the payments period, and the demand for money. *Journal of Political Economy 78*, 1228–1263.

Bollerslev, T. and I. Domowitz (1993). Trading patterns and prices in the interbank foreign exchange market. *The Journal of Finance 48*, 1421–1443.

Brock, W. A. and A. W. Kleidon (1992). Periodic market closure and trading volume: A model of intraday bids and asks. *Journal of Economic Dynamics and Control 16*, 451–489.

Burns, A. F. and W. C. Mitchell (1946). *Measuring Business Cycles*. New York: Columbia Univ. Press (for NBER).

Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica 41*, 135–156.

Dacorogna, M. M., R. Gençay, U. Müller, R. B. Olsen, and O. V. Pictet (2001). *An Introduction to High-Frequency Finance*. San Diego, California: Academic Press.

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

Feller, W. E. (1971). *An Introduction to Probability Theory and its Applications*. (2nd ed.). New York, NY: John Wiley and Sons.

Gallant, R. A., P. Rossi, and G. E. Tauchen (1992). Stock prices and volume. *The Review of Financial Studies 5*, 199–242.

Geman, H. and T. Ané (1996). Stochastic subordination. *Risk 9*(9), 145–149.

Geman, H., D. B. Madan, and M. Yor (2000). Asset prices are Brownian motion: only in business time. In M. Avellaneda (Ed.), *Quantitative Analysis in Financial Markets*. World Scientific.

Harris, L. (1982). *A theoretical and empirical analysis of the distribution of speculative prices and of the relation between absolute price change and volume*. Ph. D. thesis, University of Chicago.

Jones, C. M., G. Kaul, and M. L. Lipson (1994). Transactions, volume, and volatility. *The Review of Financial Studies 7*(4), 631–7651.

Karpoff, J. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis 22*, 109–126.

Monroe, I. (1978). Processes that can be embedded in brownian motion. *The Annals of Probability 6*(1), 42–56.

Müller, U. A. (1991). Specially weighted moving averages with repeated application of the ema operator. Internal document UAM. 1991-10-14.

Murphy, A. and M. Izzeldin (2006). Order flow, transaction clock, and normality of asset returns: A comment on Ané and Geman (2000). Working Paper, Lancaster University Management School.

Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance 44*, 1115–1153.

Tauchen, G. and M. Pitts (1983). The price variability–volume relationship on speculative markets. *Econometrica 51*, 485–505.

Velasco–Fuentes, R. and K. Chourdakis (2007). Trading activity, operational time and the normality of asset returns. Working Paper.

Wasserfallen, W. and H. Zimmerman (1985). The behavior of intradaily exchange rates. *Journal of Banking and Finance 9*, 55–72.

Zumbach, G. and U. Müller (2001). Operators on inhomogeneous time series. *International Journal of Theoretical and Applied Finance 4*(1), 147–177.

A Intraday patterns for trades and volume

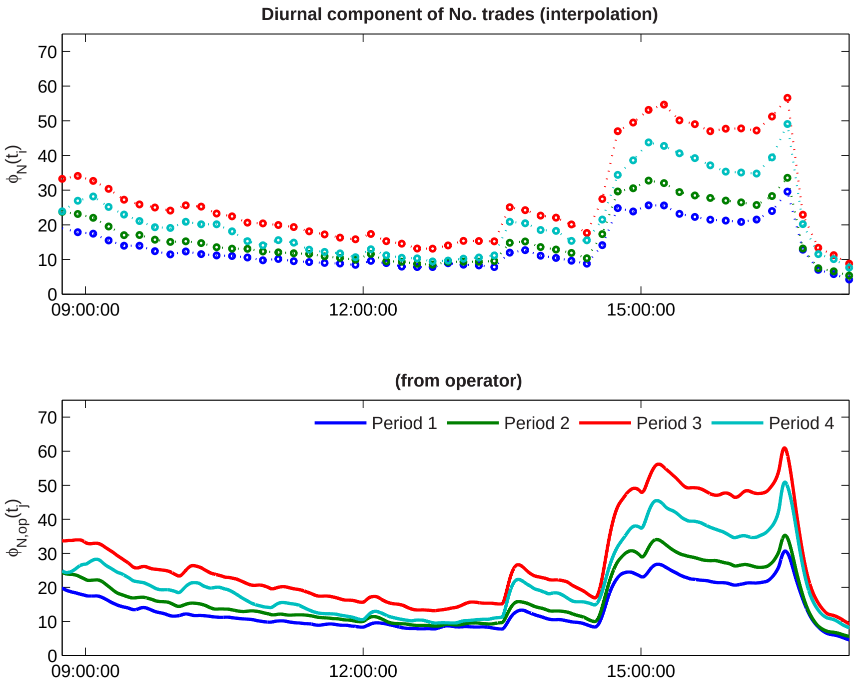


Figure 6: Diurnal component of the (cumulative) number of trades. Top: The deterministic component for the four different sub-intervals as computed by counting the number of trades during 10 minute intervals. Bottom: The same components computed using the differential operator. In both cases the number of trades have been divided by 10 in order to show number of trades per minute. The diurnal components show that although the overall level of trading appears to have increased throughout the four sub–periods, the third sub-period (red) had the highest activity measured by the number of trades. The sudden increase in activity when the US markets open can also be observed.

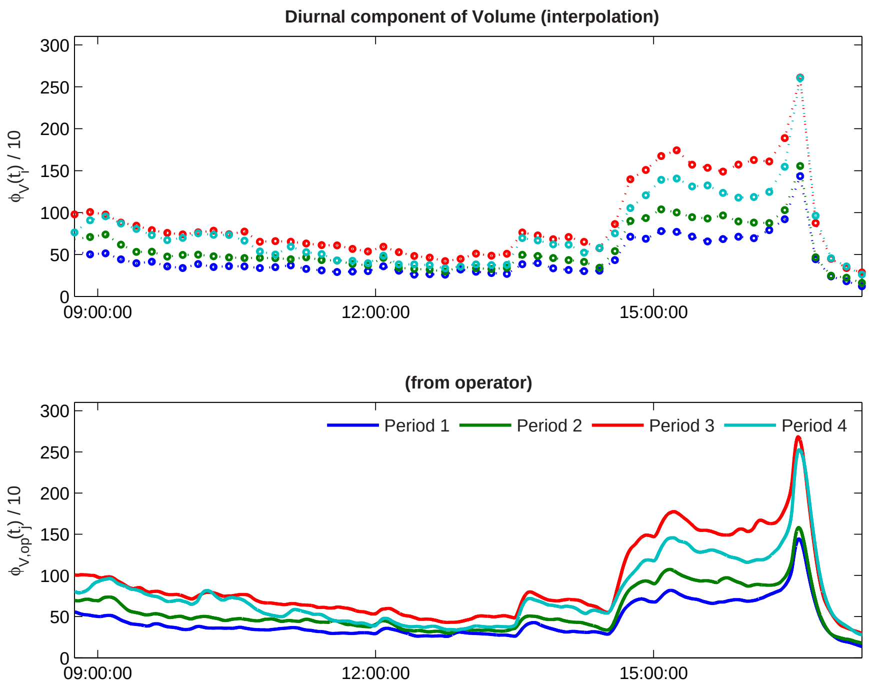


Figure 7: Diurnal component of the (cumulative) volume. Top: The deterministic component for the four different sub-intervals computed by observing the volume traded during each 10 minute interval. Bottom: The same components computed using the differential operator. In both cases the (cumulative) volume has been divided by 10 in order to show volume per minute. The diurnal components show that although the overall level of traded volume appears to have increased throughout the four sub-periods, the third sub-period (red) had the highest activity measured by volume traded. The sudden increase in activity when the US markets opens can also be observed.

B Descriptive statistics and estimated functions

	Interpolation			Differential operator		
	$r(t_i)$	$N'(t_i)$	$V'(t_i)$	$r_{\mathrm{op}}(t_i)$	$N'_{\mathrm{op}}(t_i)$	$V'_{\mathrm{op}}(t_i)$
Mean	-0.72×10^{-5}	209.6658	708.0336	-2.30×10^{-5}	211.3711	714.3638
Std. Dev.	0.0015	163.9125	669.7818	0.0013	160.7792	641.4822
Skewness	-0.3980	1.7430	3.1575	-0.2963	1.7225	2.7640
Kurtosis	19.0380	7.2247	24.9928	16.7210	7.1448	17.1376
Max	0.0200	1,828.9994	15,840.4286	0.0205	1,755.0532	9,529.9379
Min	-0.0338	4.6957	8.2570	-0.0193	6.8972	14.5178

Table 1: Descriptive statistics of calendar returns, (cumulative) number of trades and (cumulative) volume at intervals of 10 minutes. Computed using an usual interpolation scheme (left) and via the differential operator (right).

Calendar Returns

	Interpolation ($r(t_i)$)				Differential operator ($r_{\mathrm{op}}(t_i)$)			
	Pre 9/11	From 9/11	Pre Jun03	From Jun03	Pre 9/11	From 9/11	Pre Jun03	From Jun03
Mean ($\times 10^{-6}$)	3.803011	-1.090322	-22.574929	-2.393440	-13.821115	-15.225895	-36.019397	-20.367028
Std. Dev.	0.001305	0.001684	0.002051	0.000862	0.001113	0.001429	0.001743	0.000740
Skewness	-0.085892	-1.464976	-0.018727	-0.152182	-0.200489	-0.550542	-0.166507	-0.212597
Kurtosis	5.876144	38.687907	8.379605	7.157584	6.264970	30.369515	8.364855	5.909175
JB-stat	2,803.16	433,156.29	16,045.40	9,634.10	3,654.01	253,463.67	9,761.12	2,919.94

Calendar Returns with intraday seasonality removed

	Interpolation ($R(t_i)$)				Interpolation ($R_{\mathrm{op}}(t_i)$)			
	Pre 9/11	From 9/11	Pre Jun03	From Jun03	Pre 9/11	From 9/11	Pre Jun03	From Jun03
Mean	0.009782	-0.007678	-0.018169	-0.002347	-0.001906	-0.021176	-0.028386	-0.023750
Std. Dev.	1.341379	1.486049	1.380536	1.341105	1.345220	1.501574	1.382577	1.331654
Skewness	-0.091493	-0.866832	-0.059512	-0.149167	-0.057632	-0.686545	-0.171841	-0.063904
Kurtosis	5.549996	20.908859	6.381069	7.514369	5.901797	21.505523	6.732540	5.268125
JB-stat	2,206.79	109,358.92	6,345.00	11,347.55	2,847.74	116,320.29	7,788.85	2,860.49

Table 2: Left: Descriptive statistics of the 10 minute calendar returns. Computed using an usual interpolation scheme ($r(t_i)$) and via the differential operator ($r_{\mathrm{op}}(t_i)$). Right: Descriptive statistics of the 10 minute calendar returns where the intraday volatility component has been removed. In almost all cases taking into account the diurnal component reduces the skewness and kurtosis considerably.

Returns conditional on $f_{\text{lin}}(\cdot)$

	Interpolation				Differential operator			
	Pre 9/11	From 9/11	Pre Jun03	From Jun03	Pre 9/11	From 9/11	Pre Jun03	From Jun03
	R conditional on $f_{\text{lin}}(N)$				R_{op} conditional on $f_{\text{lin}}(N_{\text{op}})$			
Skewness	0.047328	-0.236581	0.044639	0.018350	0.035632	-0.220094	0.008566	0.008758
Kurtosis	3.107698	7.012718	3.641746	3.252282	3.207354	8.050098	3.655917	3.015914
JB-stat	**6.89**	5,513.40	232.46	35.91	16.14	8,678.88	238.40	**0.30**
	r_V				R_V			
Skewness	0.031300	-0.035569	0.020295	0.018662	0.018302	-0.022263	-0.002469	0.021748
Kurtosis	3.169885	4.689243	3.412327	3.510802	3.152645	5.468831	3.421080	3.188837
JB-stat	10.99	964.89	94.98	145.20	**8.25**	2,058.57	98.12	20.73

Returns conditional on $f_{\text{quad}}(\cdot)$

	Interpolation				Differential operator			
	Pre 9/11	From 9/11	Pre Jun03	From Jun03	Pre 9/11	From 9/11	Pre Jun03	From Jun03
	R conditional on $f_{\text{quad}}(N)$				R_{op} conditional on $f_{\text{quad}}(N)$			
Skewness	0.054697	-0.013147	0.066627	0.040722	0.039891	-0.061577	0.051151	0.001804
Kurtosis	3.026455	4.631357	3.427233	3.069849	3.014081	5.302974	3.368337	3.001463
JB-stat	**5.16**	898.50	110.85	**6.35**	**2.21**	1,795.74	80.85	**0.01**
	r conditional on $f_{\text{quad}}(V)$				R conditional on $f_{\text{quad}}(V)$			
Skewness	0.037693	0.092628	0.023556	0.018777	0.026505	0.096622	0.013065	0.018164
Kurtosis	3.137762	3.704902	3.341060	3.465914	3.025420	3.844405	3.304826	3.161043
JB-stat	**8.26**	179.11	65.56	120.92	**1.16**	253.08	51.75	15.03

Table 3: Skewness, kurtosis and JB statistic of the conditional returns. Left: Returns conditional on the (re-centered) number of trades, and conditional on the (re-centered) volume. Right: Returns conditional on a quadratic function of the number of trades $f_{\text{quad}}(N)$ and a quadratic function of the volume $f_{\text{quad}}(V)$. With lower skewness and kurtosis much closer to 3 for all subperiods (compared with the returns in Table 2), conditional returns are much closer to Gaussian and in some cases (bold) their JB statistics are lower than the critical value at 1% significance level.

Returns conditional on $f^{\mp}_{\mathrm{lin}}(\cdot)$

	Interpolation				Differential operator			
	Pre 9/11	From 9/11	Pre Jun03	From Jun03	Pre 9/11	From 9/11	Pre Jun03	From Jun03
	R conditional on $f^{\mp}_{\mathrm{lin}}(N)$				R_{op} conditional on $f^{\mp}_{\mathrm{lin}}(N)$			
Skewness	-0.001766	0.395247	-0.022408	0.006054	0.000048	0.225223	-0.006843	0.000001
Kurtosis	3.091823	6.090273	3.625672	3.246168	3.167362	7.645563	3.651175	3.000451
JB-stat	**2.81**	3,435.82	217.87	33.56	9.38	7,357.10	234.90	**0.00**
	R conditional on $f^{\mp}_{\mathrm{lin}}(V)$				R_{op} conditional on $f^{\mp}_{\mathrm{lin}}(V)$			
Skewness	-0.002388	-0.009338	-0.011915	0.007738	-0.000473	-0.023294	-0.004176	-0.003154
Kurtosis	3.162993	4.668503	3.407635	3.479442	3.081305	5.466298	3.419589	3.177474
JB-stat	**8.90**	939.78	92.25	127.35	**2.19**	2,054.41	97.45	17.40

Returns conditional on $f^{\mp}_{\mathrm{quad}}(\cdot)$

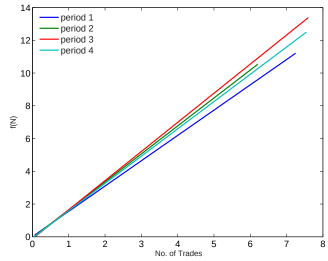
	R conditional on $f^{\mp}_{\mathrm{quad}}(N)$				R_{op} conditional on $f^{\mp}_{\mathrm{quad}}(N)$			
Skewness	-0.000000	0.134575	-0.015173	-0.000476	0.000000	0.191894	-0.009489	-0.000000
Kurtosis	3.000740	4.297357	3.403183	3.021105	3.000740	4.760537	3.359413	3.000451
JB-stat	**0.00**	592.45	90.45	**0.24**	**0.00**	1096.00	71.65	**0.00**
	R conditional on $f^{\mp}_{\mathrm{quad}}(V)$				R_{op} conditional on $f^{\mp}_{\mathrm{quad}}(V)$			
Skewness	-0.002464	-0.011923	-0.006319	0.006595	-0.000000	-0.058793	-0.004853	-0.001535
Kurtosis	3.129211	3.672517	3.330182	3.428235	3.000740	3.737143	3.301704	3.143611
JB-stat	**5.58**	152.65	60.38	101.57	**0.00**	187.88	50.37	11.37

Returns conditional on $f_{\mathrm{quad}}(N, V)$

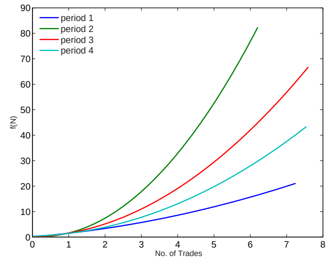
	Interpolation				Differential operator			
	Pre 9/11	From 9/11	Pre Jun03	From Jun03	Pre 9/11	From 9/11	Pre Jun03	From Jun03
	R conditional on $f_{\mathrm{quad}}(N, V)$				R_{op} conditional on $f_{\mathrm{quad}}(N, V)$			
Skewness	0.037176	0.064983	0.035139	0.036060	0.024382	0.060390	0.024654	0.001032
Kurtosis	3.046543	3.382584	3.295189	3.054910	3.013802	3.480070	3.268560	3.000994
JB-stat	**2.58**	54.96	50.91	**4.53**	**0.86**	82.55	41.21	**0.00**

Table 4: Skewness, kurtosis and JB statistic of the conditional returns. Left: Returns conditional via linear function $f^{\mp}_{\mathrm{lin}}(\cdot)$. Middle: Returns conditional via quadratic function $f^{\mp}_{\mathrm{quad}}(\cdot)$. Right: Returns conditional on $f_{\mathrm{quad}}(N, V)$ – a quadratic function of the trades *and* volume. The conditional returns with JB statistics lower than the critical value at 1% significance level are shown in bold.

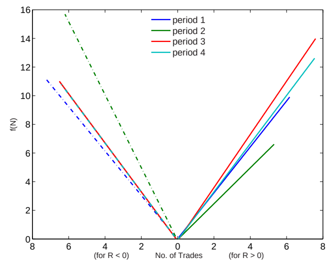
$$f(N) = 1.54N - .007$$
$$f(N) = 1.71N - .006$$
$$f(N) = 1.78N - .127$$
$$f(N) = 1.67N - .076$$



$$f(N) = .240N^2 + 1.16N + .106$$
$$f(N) = 2.31N^2 - 1.10N + .399$$
$$f(N) = 1.12N^2 + .239N + .187$$
$$f(N) = .713N^2 + .310N + .419$$



$$\begin{aligned} f(N) &= (1.55N - .050)\mathbf{1}_{\{R<0\}} \\ &\quad +(1.60N - .015)\mathbf{1}_{\{R\geq 0\}} \\ f(N) &= (2.55N - .114)\mathbf{1}_{\{R<0\}} \\ &\quad +(1.25N - .042)\mathbf{1}_{\{R\geq 0\}} \\ f(N) &= (1.71N - .131)\mathbf{1}_{\{R<0\}} \\ &\quad +(1.86N - .126)\mathbf{1}_{\{R\geq 0\}} \\ f(N) &= (1.70N - .119)\mathbf{1}_{\{R<0\}} \\ &\quad +(1.68N - .075)\mathbf{1}_{\{R\geq 0\}} \end{aligned}$$



$$\begin{aligned} f(N) &= (.463N^2 + .966N + .001)\mathbf{1}_{\{R<0\}} \\ &\quad +(.750N^2 + .095N + .631)\mathbf{1}_{\{R\geq 0\}} \\ f(N) &= (2.94N^2 - 1.11N + .232)\mathbf{1}_{\{R<0\}} \\ &\quad +(1.72N^2 - 1.12N + .586)\mathbf{1}_{\{R\geq 0\}} \\ f(N) &= (.935N^2 + .278N + .188)\mathbf{1}_{\{R<0\}} \\ &\quad +(1.39N^2 + .000N + .225)\mathbf{1}_{\{R\geq 0\}} \\ f(N) &= (1.22N^2 - .234N + .399)\mathbf{1}_{\{R<0\}} \\ &\quad +(.543N^2 + .737N + .263)\mathbf{1}_{\{R\geq 0\}} \end{aligned}$$

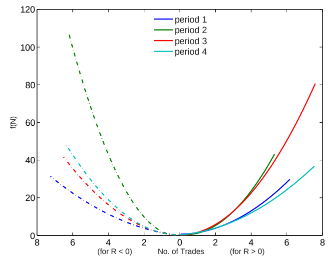
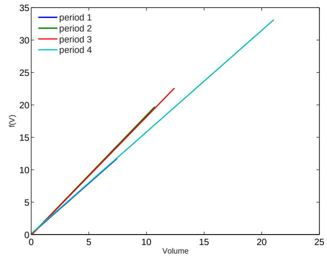
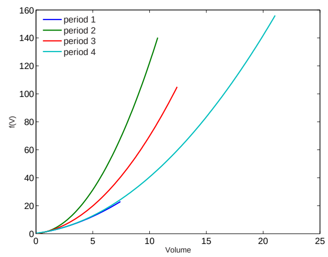


Table 5: Estimated function (and parameters) of the number of trades, where the returns and number of trades have been computed by usual interpolation sampling.

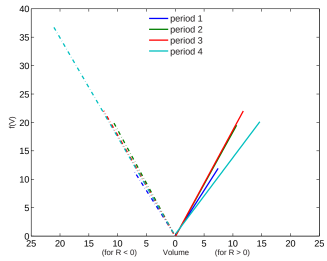
$$f(V) = 1.57V + .083$$
$$f(V) = 1.84V - .065$$
$$f(V) = 1.82V - .081$$
$$f(V) = 1.57V + .15$$



$$f(V) = .255V^2 + 1.17V + .172$$
$$f(V) = 1.19V^2 + .266V + .197$$
$$f(V) = .604V^2 + .934V + .115$$
$$f(V) = .305V^2 + .965V + .360$$



$$f(V) = (1.60V + .025)\mathbf{1}_{\{R<0\}} + (1.59V + .101)\mathbf{1}_{\{R\geq 0\}}$$
$$f(V) = (1.87V - .066)\mathbf{1}_{\{R<0\}} + (1.84V - .077)\mathbf{1}_{\{R\geq 0\}}$$
$$f(V) = (1.78V - .079)\mathbf{1}_{\{R<0\}} + (1.87V - .086)\mathbf{1}_{\{R\geq 0\}}$$
$$f(V) = (1.74V + .006)\mathbf{1}_{\{R<0\}} + (1.35V + .323)\mathbf{1}_{\{R\geq 0\}}$$



$$f(V) = (.399V^2 + .944V + .189)\mathbf{1}_{\{R<0\}} + (.197V^2 + 1.31V + .159)\mathbf{1}_{\{R\geq 0\}}$$
$$f(V) = (1.51V^2 - .581V + .483)\mathbf{1}_{\{R<0\}} + (1.08V^2 + .591V + .131)\mathbf{1}_{\{R\geq 0\}}$$
$$f(V) = (.364V^2 + 1.25V + .031)\mathbf{1}_{\{R<0\}} + (.909V^2 + .579V + .196)\mathbf{1}_{\{R\geq 0\}}$$
$$f(V) = (.364V^2 + 1.19V + .135)\mathbf{1}_{\{R<0\}} + (.323V^2 + .643V + .606)\mathbf{1}_{\{R\geq 0\}}$$

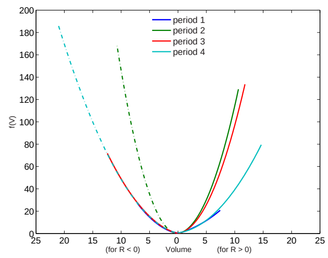
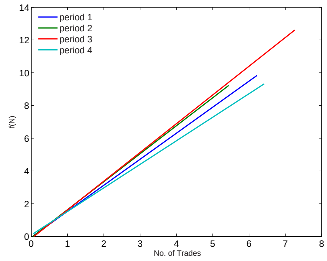
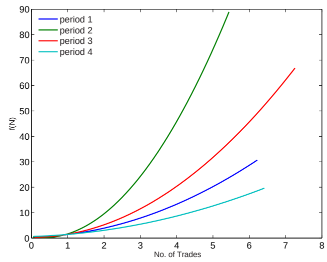


Table 6: Estimated function (and parameters) of the volume, where the returns and volume have been computed by usual interpolation sampling.

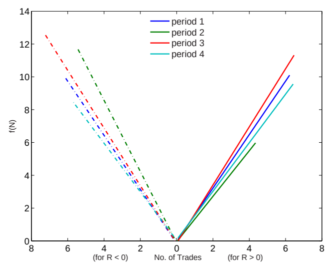
$$f(N) = 1.59N - 0.05$$
$$f(N) = 1.71N - 0.08$$
$$f(N) = 1.75N - 0.13$$
$$f(N) = 1.44N + 0.09$$



$$f(N) = .738N^2 + .277N + .391$$
$$f(N) = 3.42N^2 - 2.36N + .583$$
$$f(N) = 1.28N^2 - .099N + .301$$
$$f(N) = .395N^2 + .431N + .602$$



$$f(N) = (1.65N - .158)\mathbf{1}_{\{R<0\}} + (1.63N - .039)\mathbf{1}_{\{R\geq 0\}}$$
$$f(N) = (2.17N - .112)\mathbf{1}_{\{R<0\}} + (1.39N - .061)\mathbf{1}_{\{R\geq 0\}}$$
$$f(N) = (1.76N - .151)\mathbf{1}_{\{R<0\}} + (1.77N - .133)\mathbf{1}_{\{R\geq 0\}}$$
$$f(N) = (1.48N + .034)\mathbf{1}_{\{R<0\}} + (1.48N + .064)\mathbf{1}_{\{R\geq 0\}}$$



$$f(N) = (.600N^2 + .223N + .347)\mathbf{1}_{\{R<0\}} + (.820N^2 + .117N + .304)\mathbf{1}_{\{R\geq 0\}}$$
$$f(N) = (4.88N^2 - 2.99N + .591)\mathbf{1}_{\{R<0\}} + (2.49N^2 - 1.96N + .646)\mathbf{1}_{\{R\geq 0\}}$$
$$f(N) = (1.18N^2 + .000N + .248)\mathbf{1}_{\{R<0\}} + (1.22N^2 + .132N + .225)\mathbf{1}_{\{R\geq 0\}}$$
$$f(N) = (.207N^2 + .421N + .350)\mathbf{1}_{\{R<0\}} + (.126N^2 + .671N + .197)\mathbf{1}_{\{R\geq 0\}}$$

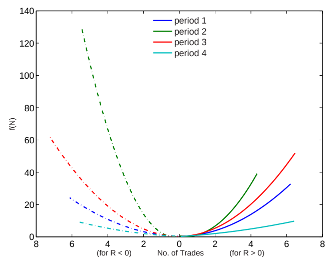
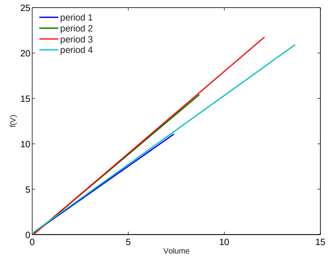
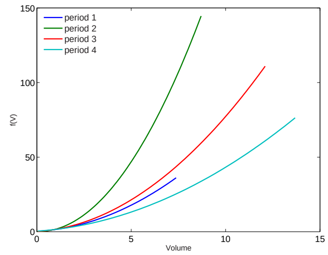


Table 7: Estimated function (and parameters) of the number of trades, where the returns and number of trades have been computed by sampling of the operator.

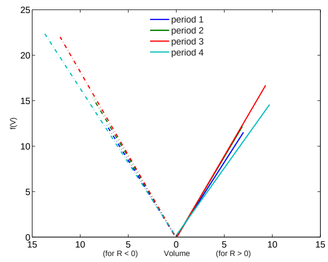
$$
\begin{aligned}
f(V) &= 1.49V + 0.10 \\
f(V) &= 1.78V - 0.07 \\
f(V) &= 1.81V - 0.11 \\
f(V) &= 1.52V - 0.14
\end{aligned}
$$



$$
\begin{aligned}
f(V) &= .570V^2 + .656V + .273 \\
f(V) &= 1.96V^2 - .471V + .222 \\
f(V) &= .692V^2 + .794V + .114 \\
f(V) &= .344V^2 + .844V + .381
\end{aligned}
$$



$$
\begin{aligned}
f(V) = (1.71V - .114)\mathbf{1}_{\{R<0\}} \\
+(1.65V + .004)\mathbf{1}_{\{R\geq 0\}} \\
f(V) = (1.77V - .070)\mathbf{1}_{\{R<0\}} \\
+(1.78V - .074)\mathbf{1}_{\{R\geq 0\}} \\
f(V) = (1.83V - .129)\mathbf{1}_{\{R<0\}} \\
+(1.81V - .110)\mathbf{1}_{\{R\geq 0\}} \\
f(V) = (1.63V + .013)\mathbf{1}_{\{R<0\}} \\
+(1.48V + .197)\mathbf{1}_{\{R\geq 0\}}
\end{aligned}
$$



$$
\begin{aligned}
f(V) = (.895V^2 + .448V + .173)\mathbf{1}_{\{R<0\}} \\
+(.421V^2 + 1.12V + .072)\mathbf{1}_{\{R\geq 0\}} \\
f(V) = (2.04V^2 - 1.11V + .473)\mathbf{1}_{\{R<0\}} \\
+(2.38V^2 - .561V + .208)\mathbf{1}_{\{R\geq 0\}} \\
f(V) = (.792V^2 + .596V + .173)\mathbf{1}_{\{R<0\}} \\
+(.619V^2 + .972V + .055)\mathbf{1}_{\{R\geq 0\}} \\
f(V) = (.227V^2 + 1.27V + .113)\mathbf{1}_{\{R<0\}} \\
+(.501V^2 + .440V + .601)\mathbf{1}_{\{R\geq 0\}}
\end{aligned}
$$

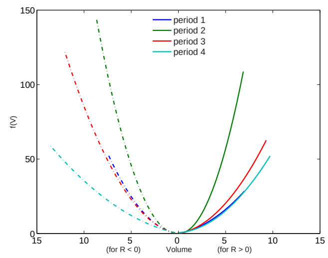


Table 8: Estimated function (and parameters) of the volume, where the returns and volume have been computed by sampling of the operator.