TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

SEF Working paper: 05/2014

On long memory behaviour and predictability of financial markets

Long H. Vo and Leigh Roberts

Further enquiries to:

The Administrator
School of Economics and Finance
Victoria University of Wellington
P O Box 600
Wellington 6140
New Zealand

Phone: +64 4 463 5353
Email: alice.fong@vuw.ac.nz

Working Paper 05/2014

ISSN 2230-259X (Print)

ISSN 2230-2603 (Online)

On long Memory Behaviour and Predictability of Financial Markets

Long H. Vo* and Leigh Roberts†

May 8, 2014

Abstract

An immediate consequence of the Efficient Market Hypothesis (EMH) is the absence of autocorrelation of the return series of the financial prices and the exclusion of excess profitability made by any (active) trading strategy. However, the precondition for the validity of EMH, which assumes that all market participants can promptly receive and rationally react to the relevant information affecting the prices, might be (approximately) true for a long time horizon, but not for a short time horizon. By examining local long-range dependence (measured by the rolling Rescaled Range estimates of the Hurst index) of an empirical example, the local market inefficiency is inferred, and excess profitability of a simple trend-following trading strategy is observed. Moreover, the significant positive cross-correlation between the local Hurst index estimates and the returns of the trend-following trading strategies implies the potential for constructing a more profitable trading system by incorporating the former into the latter.

Key words: High-frequency trading, Hurst index, Long memory, Market efficiency, Rescaled range analysis, Trading system

1 Introduction

Seeking excess profit is the ultimate goal (or dream) for many financial market participants. This goal is deemed unachievable by many researchers and practitioners, based on the perception of the classic Efficient Market Hypothesis (EMH), which was independently proposed by (Samuelson, 1965) and (Fama, 1965). However, this hypothesis, which claims that the market can instantly and correctly react to all price-related events, is rather idealistic than realistic. Intuitively, the dissemination of price-related information may take days or weeks, or even months, and the market reaction to a particular event is often biased at the beginning. Indeed, it would be unreasonable to assume that efficiency can be maintained consistently, as evidenced by numerous incidents leading to inefficiencies such as value stocks and small firms yielding returns higher than the market average, or the various crashes over the years implying severely mispriced assets. Therefore, many modern financial economists (e.g. Lo, 2004) reject

*Corresponding author. School of Economics and Finance, Victoria University of Wellington.
E-mail: longvoqnu@gmail.com

†School of Economics and Finance, Victoria University of Wellington.
E-mail: leigh.roberts@vuw.ac.nz

the notion of a "static" sense of market efficiency, and adopt an adaptive, evolutionary perspective instead. In other words, locally or over a short time horizon, the market exhibits various degrees of efficiency.

It is well-known that the EMH is closely related to the random walk hypothesis (RWH) which proposes that the financial returns (usually defined as differences of logarithm of the prices) can be modelled by a random walk process, and martingale theory which assumes that the return series is a martingale process. Both theories suggest that the consecutive financial returns in an efficient market are uncorrelated. On the other hand, the returns in an inefficient market are expected to be correlated to, or rather dependent on, each other to some degree. Therefore, the (serial) dependence, in particular the long range dependence (hereafter LRD) or "long memory", of the financial returns becomes an indicator of market inefficiency.

A stylized fact well-documented in financial time series literature is the absence of significant autocorrelations of returns. This makes the notion of serial correlation (autocorrelation) inappropriate when describing the LRD of the financial returns. However, according to (Cont, 2005), quoted in (Heyde, 2002), one should distinguish dependence in sign from that in amplitude of the returns (measured by squared or absolute returns, which is analogous to volatility). It is the latter rather than the former that account for the (long range) dependence, and moreover the predictability of the returns series. In other words, the autocorrelation structure of the volatility series is more informative in terms of describing the long range dependence of financial time series and this was supported by the study of another stylized fact called "volatility clustering".

From a time series perspective, the two prominent and relating features exhibited by the volatility of returns are clustering and long-memory. The phrase 'clustering' is used to describe the situation when large/small price fluctuations are most likely followed by a movement similar in magnitude. From a different angle, 'long-memory' implies the fact that past volatility fluctuations can have significant impact on the behaviour of returns series in the distant future. In other words, past returns 'clusters' are strongly related to future 'clusters'. In addition, long-memory is visually reflected via the slowly decaying correlation between subsequent observations of absolute returns (or squared returns). Another way to think of this persistent behaviour is proposed by (Engle and Patton, 2001), who point out that present returns tend to have a large impact on future volatility forecasts. This dependence structure will be our primary focus. All in all, these well-established stylized facts generally imply a certain degree of predictability of volatility (and returns in the short run), as opposed to randomly generated returns which are uncorrelated (in the long run). Furthermore, we proceed to emphasize the link between some of the most popular trading strategies based on technical analysis and the predictability implied by the (local) LRD behaviour of stock returns.

With these objectives in mind, the rest of the paper is structured as follows: section 2 revisits the 'conflict' between LRD and efficiency in financial markets. Section 3 reviews the application of a technique, known as the Rescaled range analysis, that aims to estimate the LRD parameter. The first two sections serve as a theoretical background for later empirical studies, which focus on one specific company in the financial service industry, Citigroup Inc., mainly for its central role in the recent financial crisis. In section 4 we provide the summary statistics of our daily data as well as preliminary data analyses. Section 5 mainly revolves around the long-memory property of the actual data generating processes. Most notably, we document a connection between the time varying nature of LRD and the trending behaviour of the stock market. It is suggested that this relationship might be a good indicator of profitable trading rules. Section 6 offers some concluding remarks and comments on the general implications of our study, as well as opening up possible venues for future research.

2 Long range dependence of financial returns and market efficiency

In stock markets, analyses of long-range dependence of returns are known to yield mixed evidence: for example, in the long run stock indices are observed to have displayed long-memory by (Mandelbrot, B., 1971), while contrasting evidence is reported by (Lo, 1991). The implications of these studies create a focal point for intensive debate. This is because the existence of long-memory generally indicates predictability of future returns based on past returns, which violates the basic assumption of one of the most strongly supported ideas in the history of economics, the Efficient Market Hypothesis (EMH). The EMH (Samuelson, 1965; Fama, 1965) in its strongest form, generally assumes that the changes of stock price follow a random walk. The intuition (or seemingly counter-intuition!) is, when all available information and/or all expectation is fully reflected in prices, one cannot forecast the price changes by simply looking at past prices. Ironically, any informative advantage, even the smallest, is instantaneously exposed and incorporated into market prices when the investors possessing it try to make profit from it. Therefore classic EMH implies instant access to information. In this 'ideal' scenario, prices are also said to follow a martingale, which is the cornerstone of traditional asset pricing and derivative pricing models. Therefore, violation of this condition would undermine the foundation of these models. For example, linear models of returns such as the classic Capital Asset Pricing Model (CAPM) will encounter numerous problems should price changes not be random. Furthermore, if long-range dependence exists, implications from economics disciplines that are sensitive to investment horizons such as optimal consumption decision and portfolio management would be affected (Lo, 1991).

In contrast to the mixed evidence of long-memory in returns, such behaviour is widely observed to be a 'stylized fact' of the volatility of financial returns. (Ding et al., 1993), (Andersen and Bollerslev, 1997) are among the advocates of this vein of thought, while (Lo, 1991) opposed it. In any case, widely documented long range dependence displayed by time series from multiple economics contexts has inspired (Mandelbrot and Wallis, 1968) to relate this phenomenon to the prophecy made by Joseph (in a biblical reference from the Old Testament), who predicted that Egypt was to have seven years of prosperity followed by seven years of famine. Hence the fanciful yet perhaps aptly termed "Joseph effect", often accompanies the more well-known "Hurst effect" in the long-memory literature.

Some academics consider the Hurst exponent, or the "index of dependence", as a component of the so-called Chaos Theory (see e.g. (Peters, 1996)). Based on this theory, an alternative to the EMH, the Fractal Market Hypothesis (FMH), is proposed. This hypothesis casts doubt on the ideally uniform and simultaneous interpretation of information reflected in prices, as embraced by the EMH. Instead it assumes that traders may decipher information in different ways and at different times. If investors were influenced by events unfolding from the past, price changes might not be entirely unpredictable, and the Hurst index might be different from 0.5. The FMH also assumes non-normal, leptokurtic distribution of price changes and prices decreasing faster than they increase, all of which are empirically true. Intuitively, if stock prices follow a random walk/Brownian motion under the general assumption of the EMH, then their logarithmic differences (or the financial returns, as we shall discuss in the next section) should be normally distributed. Yet in practice the overwhelming evidence of heavy tailed returns distribution suggests stock prices do exhibit dependence to some extent, thus invalidating the EMH. Perhaps one of the strongest criticism against the EMH that caught widespread attention of academics to date is presented in the book of (Lo and MacKinlay, 1999). More directly relevant to our study, it is the stylised clustering behaviour of stock returns and the predictability of the financial data generating process due to its inherent long-memory that have shaken the universal foundation of market efficiency.

These observations have led to a new consensus that relates efficiency to economic development. In particular, as economies gradually evolve from a under-developed to a sophisticated state we would expect to see a corresponding movement towards efficiency of financial markets in the form of correctly priced assets. Obviously this is not a static process, nor is it a short-term one. Adopting this approach, (Hull and McGroarty, 2013) use a Hurst index estimate to show that different stages of emerging economies do correspond to increasing levels of efficiency and exhibit different degrees of long-memory. The principle of this idea is consistent with (Grossman and Stiglitz, 1980) who support a ‘self-correction' viewpoint in which arbitrageurs attracted by mispriced assets would eventually enforce efficiency and thus, in a way, inefficiency takes an indispensable role in maintaining efficiency itself. Along the same vein of thought, in his path-breaking article, (Lo, 2004) attempted to reconcile the assumption of market rationality with the various psychological aspects of the documented irrationality among investors, and introduced another alternative to the EMH, the Adaptive Market Hypothesis (AMH). In essence, the AMH is a synthetic compromise between two seemingly conflicting schools of thinking: the EMH and Behavioural finance. The latter advocates ubiquitous behavioural biases (e.g. overconfidence, over-reaction or herding) that could lead to distortions of utility optimising decisions that form the basis of the former. In a sense, this means that to the AMH, extreme market movements such as crashes are nothing more than conditions facilitating a ‘natural selection' process that casts out investors that could not adapt to the ever changing market environment. As such, compared to the EMH, the AMH implies “*[C]onsiderably more complex market dynamics, with cycles as well as trends, and panics, manias, bubbles, crashes, and other phenomena that are routinely witnessed in natural market ecologies.*” ((Lo, 2004), p.24).

It is observed that equity volatility exhibits a Hurst exponent estimated to be greater than 0.5, typically being 0.7 (Peters, 1996). It would be interesting then, to reconcile the trending behaviour of stock returns implied by Hurst index estimates and the trend-detecting techniques which form the basis of so-called “technical analysis”. As it turns out, using a trading rule designed for capitalising the trending behaviour of stock price during certain periods, (Mitra, 2012) documented greater trading profit associated with higher long-memory parameter during the periods studied. On the other hand this author also observed lower profits at times when the market exhibits mean-reverting behaviour. Intuitively, technical trading strategies, which follows possible market ‘trends’, are expected to exhibit some correlation with the Hurst index, which is a reasonable measure of such trending behaviour. In our empirical study in Section 5 we show that this is indeed the case.

It is surprising to find that in spite of the intensive studies on the relationship between Hurst index and the LRD of financial markets, as well as some well-established observations on the trending behaviour of markets and their predictability, there is virtually no recognised empirical research on the topic of applying the information conveyed by the Hurst index to actually forecast the market movements. One possible exception is (Xu and Jin, 2009), who use local Hurst index estimates to predict drastic crashes of a Chinese stock index, with relatively robust results.

3 LRD processes and estimating LRD parameter

We construct the current section as follows: in subsections 3.1 and 3.2, we shall cover the fundamentals of typical stationary long-memory processes to provide a theoretical background for studying our financial time series in later sections. Subsection 3.3 then demonstrates the well-established methodology of Rescaled range, which is designed to estimate the degree of long-memory characterized by the level of the Hurst exponent H. Subsequently, by separately applying this method to a data set simulated using one of the processes described, we are able to confirm that the estimator provided by Rescaled range analysis is generally robust.

3.1 General discussions of long-memory and self-similarity

The so-called ‘self-similarity parameter’ associated with long-memory process has a rich history. A definition is provide by (Dieker and Mandjes, 2003) for both discrete-time and continuous-time stochastic processes. Here we only restate the definition in the continuous context: a process $\{X(t)\}$ $(0 \leq t < \infty)$ is said to be self-similar if the two processes: $\{X(at)\}$ and $\{a^H X(t)\}$ have identical finite-dimensional distributions for all $a > 0$. The parameter a can be thought of as a scaling parameter so the latter process is actually a scaled version of the former. Analogously to the notion regarding stationarity, there exists a weaker form of self-similarity when these processes have equal mean and covariance structure, in which case we called it second-order self-similar (Cox, 1984). So what is the role of the Hurst exponent H? To be more specific, provided that the above condition holds when $0 < H < 1$, we have a self-similar process. In the special case of $1/2 < H < 1$, for a (weakly) stationary process, the second-order self-similarity also implies long range dependence among present and distant past values of that process, a feature commonly referred to as “long-memory”.

A definition of a broad long-memory class can be found in (Beran, 1994), which states that such a process is defined if its autocorrelation function $\rho(l) = \dfrac{\mathrm{Cov}(X_i, X_{i+l})}{\mathrm{Var}(X_i)}$ is non summable and satisfies $\sum_{l=0}^{\infty} \rho(l) = \infty$. In this context we assume $\{X_t\}$ to be a weakly stationary discrete time series. This implies a slowly decaying autocorrelation function, i.e.

$$\rho(l) \sim C|l|^{-\alpha} \quad \text{when} \quad |l| \to \infty$$

This is the basic property of all processes belonging to the long-memory class, where C is a constant and $0 < \alpha < 1$ is a parameter representing the decay rate (we will show later that in general $\alpha = 2 - 2H$). In this case the decay is said to follow a "power law". Larger H implies stronger long-range dependence, or more persistent impact of past events on present events. Conventional statistical inference for processes exhibiting this feature can be dramatically altered. As (Dieker and Mandjes, 2003) pointed out, for a process with finite variance and/or summable covariances such as an AR(1) process, the standard deviation of its mean is asymptotically proportional to $n^{1/2}$. This is a crucial condition for traditional statistical inference to be meaningful. However, with long-range dependence introduced by a slowly decaying $\rho(l)$, the same standard deviation is proportional to $n^{-\alpha/2}$, thus affecting all relevant test statistics, as well as the confidence intervals for the estimate of the sample mean.

The long-memory processes are contrasted with the short-memory class, which exhibits summable and exponentially decaying covariances (which is also termed short-range dependence or weak dependence). ((Lo, 1991), p.1281) made a clear distinction between these two classes, asserting that the short-memory behaviour is characterized by the fact that "*...[T]he maximal dependence between events at any two dates becomes trivially small as the time span between those two dates increases.*". In other words, the rate at which dependence decays is very high for processes exhibiting short-run dependence. Here we are only interested in what this distinction means in an empirical financial context.

3.2 Some popular long-memory data generating processes

The commonly observed phenomenon of long run dependence structure of volatility is, first and foremost, attributable to a long-memory data generating process (DGP hereafter). Evidence of this fact is often illustrated via a slow hyperbolic decay rate of the autocorrelation function of empirical time series in many physical sciences. Similarly, it is widely documented that the evolution of the volatility of financial assets' returns constitutes a long-memory stochastic process.

To be specific, this type of process is defined by a real number H and a constant C such that the process's autocorrelation is $\rho(l) = Cl^{2H-2}$ as the lag parameter $l \to \infty$. The parameter H is known as the Hurst exponent/index, named after the hydrologist H.E. Hurst, who first analysed the presence and measurement of long-memory behaviour in stochastic processes in 1951. B. Mandelbrot and his colleagues (see e.g. (Mandelbrot and Wallis, 1968) and (Mandelbrot and Van Ness, 1968)) proposed the idea that so-called "long-memory" processes can be thought of in a fractionally integrated sense. We know that a time series is said to be integrated of order one

(I(1)) if its first-order difference is stationary. A stationary series is then called an I(0). A long-memory process are then defined as a 'middle ground', i.e. I(d) where $0 < d < 1$. The fractional difference parameter d is related to the Hurst exponent by the simple equality $d = H - 0.5$. The simplest way to distinguish the three types of processes is to look at their autocorrelation function's pattern: infinite persistence (I(0)), exponential decay (I(1)) and hyperbolic decay (I(d)). In the following discussions we shall show that the general form of the autocorrelation of an I(d) process can be generalized as $\rho(l) = Cl^{-\alpha}$ for some constant C and integer α which is proportional to d and H. Generally, the interpretation of H and d with regards to the nature of long-memory is summarised in Table 1:

Hurst index	Fractional difference parameter	Behaviour of the process
H $\leq$ 0	d $\leq$ -1/2	Non stationary
0 < H < 1/2	-1/2 < d < 0	Anti-persistent, mean reverting
H = 1/2	d = 0	Uncorrelated, random process
1/2 < H < 1	0 < d < 1/2	Long range dependence
$H \geq 1$	d $\geq$ 1/2	Non stationary

Table 1: Categorizing stochastic processes based on their long-memory property.

Fractional Gaussian noise (fGn) We first introduce the process known as fractional Brownian motion (fBm). Denoted as $\{B_H(t), t \geq 0\}$, the fBm is a stochastic Gaussian process with mean zero, stationary increments, variance $E\left[B_H^2(t)\right] = t^{2H}$ and covariance:

$$\gamma(s,t) = E\left[B_H(s)B_H(t)\right] = \frac{1}{2}(s^{2H} + t^{2H} - |s-t|^{2H}) \tag{1}$$

(Carmona and Coutin, 1998) provided a brief introduction to the fBm, which was then termed a 'centered Brownian motion'. Except for a different covariance structure, the fBm is analogous to the Brownian motion. Most notably, its increments are no longer independent though are still stationary.

As our direct application is for discrete data, it is only logical to move from differencing random processes to differencing time series. The Fractional Gaussian noise where $\{X_t, t = 0, 1, 2, \dots\}$ is the first-order differenced process of the fBm, i.e.

$$X_t = B_H(t+1) - B_H(t)$$

Like the fBm, it is also a mean zero, stationary Gaussian process with autocovariance function:

$$\gamma(l) = E[X_t X_{t+l}] = \frac{1}{2}[(l+1)^{2H} - 2l^{2H} + |l-1|^{2H}] \quad \text{with } l \geq 0 \tag{2}$$

Provided that $H \neq 0.5$, function $\gamma(l)$ satisfies $\gamma(l) \sim H(2H-1)l^{2H-2}$ as $l \to \infty$. When $H = 0.5$, $\gamma(l)$ converges to zero for large l, and $\{X_t\}$ is effectively a white noise. On the other hand, when $0.5 < H < 1$ the process realizations, the X_ts, are positively correlated and display long-range dependence.

ARFIMA (p,d,q) Throughout the literature, the leading parametric long-memory model used is the Auto regressive fractionally integrated moving average (ARFIMA) model introduced by (Granger and Joyeux, 1980) and (Hosking, 1981). An ARFIMA (p,d,q) time series $\{X_t\}$ is a generalized version of the simpler fractional ARIMA (0,d,0), and can be defined by:

$$\begin{aligned} &a(L)(1-L)^d X_t = b(L)z_t \\ &z_t \sim \text{i.i.d} \;\; \mathrm{N}(0,1) \end{aligned} \tag{3}$$

where $a(L)$ and $b(L)$ are polynomials of the lag operator L of order p and q, respectively. The roots of these polynomials are assumed to lie outside the unit circle. For this type of model, we have the fractional differencing operator $(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} L^k$, which expression is known as the hyper-geometric function where the fractional parameter $-\frac{1}{2} < d < \frac{1}{2}$. Another way to express this operator is to use the Maclaurin series expansion:

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} L^k = 1 - dL - \frac{1}{2}d(d-1)L^2 - \frac{1}{6}d(d-1)(d-2)L^3 - \dots$$

When $\{X_t\}$ satisfies all of these conditions we have a stationary and invertible ARFIMA process (Bollerslev and Wright, 2000). Specifically, the process is stationary if all solutions of $a(L) = a_1 z + a_2 z^2 + \cdots + a_p z^p = 1$ are outside the unit circle $|z| = 1$ where z is a complex number. Likewise it is invertible when $b(L) = b_1 z + b_2 z^2 + \cdots + b_q z^q = -1$ has no root lying on the unit circle (Taylor, 2005).

We are able to simulate the described long-memory processes thanks to functions incorporated in software such as the fArma package in R (Wuertz, 2013). Figure 1 plots a sample path for each of these simulated processes with $H = 0.7$ and $N = 1,000$ observations. Also, we plot the estimated ACF (calculated from the samples) of the previously simulated processes and compare it to the theoretical ACF (using their respective definitions). The shape of these processes' autocorrelation functions reflect their long-range dependence structure: they decay hyperbolically at a slow rate (up to 100 lags in our example).

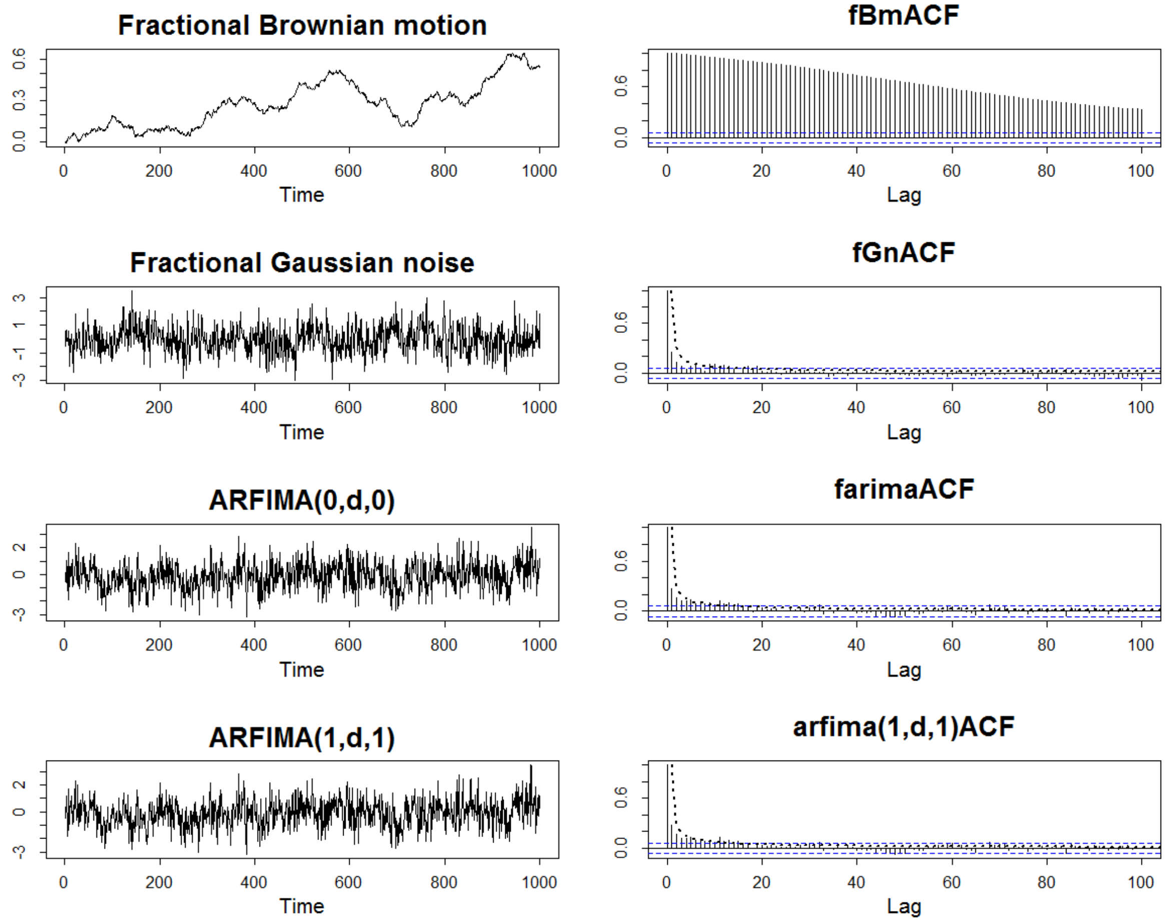


Figure 1: Left panel: sample paths of long memory processes (simulated with $H = 0.7$ ($d = 0.2$) and $N = 1000$). Right panel: their respective estimated and theoretical autocorrelation (or the approximate power curve in the general form of $\rho(l) \sim C|l|^{-\alpha}$). The theoretical ACFs are indicated by dotted curves.

3.3 Estimating long memory in financial time series with Rescaled range analysis (R/S for short)

The so-called Hurst index associated with a long-memory process has a rich history. (Hurst, 1951) was the first to proposed a method to detect and estimate the widely observed and naturally occurring empirical long-memory in the form of the "Rescaled range" statistic, denoted as $R/S(n)$ (where n represents the sample size). Assuming that the process generating the empirical data is long range dependent, this method aims to infer the Hurst index H as implied by the relationship $E[R/S(n)] \sim Cn^H$ when $n \to \infty$ and the finite positive constant C is independent of n.

Since the introduction of the R/S analysis methodology, robust empirical evidence of long range dependence in time series has been extensively documented in various disciplines, particularly from physical science studies, where studied time series exhibit some kind of trending behaviour (e.g. the length of tree rings, level of rainfall, fluctuations in air temperature, oceanic movements and volcanic activities...). Application and generalisation of the Rescaled range method were popularised by (Mandelbrot and Van Ness, 1968). Among the first to use this method to examine long-range dependence in common stock returns is (Mandelbrot, 1966). Furthermore, (Mandelbrot, 1972), together with many others, radically refined the R/S statistic. In particular, they advocate its robustness in detecting as well as estimating long-range dependence even in non-Gaussian processes with extreme degrees of skewness and kurtosis. Additionally, this method's superiority over traditional approaches such as spectral analysis or variance ratios in detecting long-memory was also shown in these researches.

However, as (Lo, 1991) pointed out, the refined statistic is not able to distinguish the effects of short-range and long-range dependence. To compensate for this weakness, he proposed a new modified R/S framework. His findings indicate that the dependence structure documented in previous studies are mostly short-ranged, corresponding to high frequency autocorrelation or heteroskedasticity. There are two important implications that we need to draw from Lo's paper: (i) empirical inferences of long-range behaviour must be carefully drawn, preferably by accounting for dependence at higher frequencies and (ii) in such cases, conventional models exhibiting short-range dependence (such as AR(1) or a random walk process) might be adequate. In the following discussion we first provide a brief description of the R/S statistic, then present an estimate of the long-range dependence parameter H for a simulated fractional Gaussian noise process.

A simple definition of the 'classic' R/S statistic is provided by (Cavalcante and Assaf, 2004) which we rearrange to our purposes: Given a series of returns $\{r_t\}$ $(t = 1, 2..., n)$ we divide it into several 'ranges', or 'blocks' with range size k satisfying $1 \leq k \leq n$, whereby the R/S statistic is:

$$Q_n = \left(\frac{R}{S}\right)_n = \frac{1}{\hat{\sigma}_n}\left[\max_{1\leq k\leq n}\sum_{t=1}^{k}(r_t - \overline{r}_n) - \min_{1\leq k\leq n}\sum_{t=1}^{k}(r_t - \overline{r}_n)\right] \tag{4}$$

Here the bracketed terms of Q_n are the maximum and minimum (over k) of the cumulative deviations of r_t from the sample mean $\overline{r}_n = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} r_t$. Because $\displaystyle\sum_{t=1}^{n}(r_t - \overline{r}_n) = 0$, the maximum term is always non-negative whereas the minimum term is always non-positive, hence the 'range' quantity R_n (the numerator of Q_n) is always non-negative, thus $Q_n \geq 0$.

We have the denominator $S_n = \hat{\sigma}_n$ as the maximum likelihood estimated standard deviation (i.e. $\hat{\sigma}_n \equiv \dfrac{1}{n}\displaystyle\sum_{t=1}^{n}[r_t - \overline{r}_n]$). In short, we 'rescale' the range of partial sums of the deviations of the time series from its mean by its standard deviation. Instead of using S_n, the modified R/S statistic proposed by (Lo, 1991) utilises the square root of a consistent estimator of the cumulative sum's variance first proposed by (Newey and West, 1987). Specifically:

$$\begin{aligned} S_n^2 &= \hat{\sigma}_n^2(l) = \hat{\sigma}_n^2 + 2\sum_{j=1}^{l} w_j(l)\ \hat{\gamma}_j \\ w_j(l) &= 1 - \frac{j}{l+1} \text{ with } l < n \end{aligned} \tag{5}$$

where $\hat{\sigma}_n^2$ and $\hat{\gamma}_j$ are the sample variance and autocovariance of order j (where $j = 1, 2, \ldots, l$). The modified denominator involves not only the sample variance of returns series, but also its weighted autocovariances up to a selected lag l. The added component can capture any short range dependencies that may appear in the data. Note that when we set the lag length l equal to zero, the modified R/S statistic reverts to its classic form.

Hurst index estimate The whole sample spreads across a time interval with time points from 1 to n. We divide this interval into u sub-intervals (where $u \in \{1, 2, ..., U\}$ with U being the integer part of k/n). Each sub-interval has length k and can be used to calculate $Q_k(u)$. Then we compute the average R/S statistic across all sub-intervals:

$$Q_u = \frac{1}{u} \sum_{u=1}^{n/k} Q_k(u) \tag{6}$$

The Hurst exponent is approximated by the slope of the regression of log Q_u against log k. The resulted plot is known as the rescaled range plot.

(Lo, 1991) noted that for short-range dependent time series, when the sample size n increases without bound, the ratio $\frac{\log Q_n}{\log n}$ "*approaches 1/2 in the limit, but converges to quantities greater or less than 1/2 according to whether there is positive or negative long-range dependence.*" (p.1289). This argument is in line with the features of the long-memory stationary processes as discussed in subsection 3.2. The procedure described above provides a visual representation of the Hurst exponent and was also called the 'graphical' R/S method. In addition, (Lo, 1991) developed a confidence interval for testing the null hypothesis of no long range dependence. Considering only the interval with length $k = n$ instead of multiple range sizes, the 95% asymptotic acceptance region for the null hypothesis ($H_0 : H = 0.5$) is that $Q_n(u) \in [0.809, 1.862]$. With this test, we are only able to detect long range dependence and still have to resort to the graphical R/S method to estimate the value of the Hurst exponent.

With regards to this modified R/S method, (Teverovsky et al., 1999) expressed concerns over the choice of the lag length l in equation 5. Given that Lo's statistic is asymptotic assuming $n \to \infty$ and $l \to \infty$ whilst in reality we only have finite sample size, what would be the appropriate choice of l? Previously we noted that as l increases, the autocorrelation $\rho(l)$ approaches Cl^{2H-2}. This means that for a typical long memory time series ($H > 0.5$), a large enough l will inevitably cause the R/S statistic to decrease and fall within the acceptance region of the null hypothesis. Application of this method to simulated time series with different degrees of dependence, as shown in (Teverovsky et al., 1999), indicates bias towards accepting the null hypothesis and illustrates the positive relationship between the R/S statistic and the lag l given $0.5 < H < 1$.

Additionally, despite the enormous praise the R/S statistic enjoyed over the years, we should also be cautious of the implication of Lo's modified method because of its tendency to reject long range dependence even when evidence of such behaviour in fact exists (albeit weakly). Another approach is suggested by (Taqqu et al., 1995): viz., the R/S method should be compared with a diverse range of well-established alternatives in the literature of LRD estimation. Nevertheless, exploring the relative performance of this technique is beyond the scope of this paper, and will be further discussed in future research.

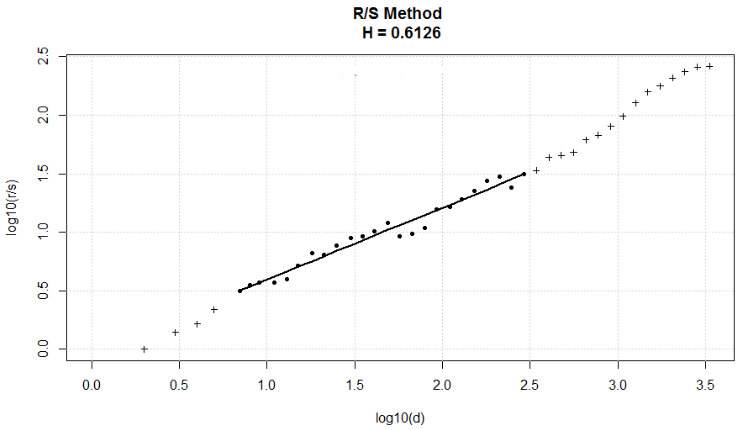


Figure 2: Log-log plot for estimating Hurst index from an fGn process (simulated with $H = 0.7$) by the (modified) R/S method.

4 Description and preliminary examination of data

4.1 Data description

Citigroup, Inc. (NYSE ticker: C) has a very rich and dynamic history, to say the least. It started off in 1812 as the City Bank of New York State. In the mid 19th century the establishment of the first transatlantic cable line provided Citigroup with a great opportunity, since the head of the telegraph firm laying the line also happened to be on Citigroup's Board of directors at the time, thus solidifying the company's initial foothold overseas. At the end of the American Civil war in 1865, Citigroup was converted to a national charter, and henceforth assumed the authority to issue U.S. bonds and currency well before the foundation of the Federal Reserve in 1913. More recently, the bank was the the first to offer travellers cheques and compound interest on deposits; the first to issue certificates of deposits and also a pioneer in adopting the modern ATMs system. These are but a very few illustration of Citigroup's innovations, all of which are closely associated with the history of the U.S. financial market. To some extent, it is not an exaggeration to say that the fluctuations this company experienced reflect the evolution of the global financial system itself. Therefore, although it can be true that one company may not provide a good indicator of a whole sector, in our opinion, with its special position Citigroup is the ideal subject to study if we are to understand the underlying mechanisms driving market movements, particularly the elusive returns-volatility dynamics. Our main focus in the next subsection would be to provide a preliminary investigation into its stock prices and returns over the last 30 years or so.

Since we generally do not have any data over weekends and holidays, we have to find a way to overcome this issue. The simplest remedy is to use concatenated returns: we use whatever data is available to compute returns over periods containing missing data associated with weekends and holidays. Obviously this method does not account for overnight returns. However, it suffices for our purposes. We collect daily closing prices of Citigroup between 03 Jan 1977 and 31 Jul 2013 from http://finance.yahoo.com and obtain a total of 9228 daily returns.

Figure 3 shows the time series of Citigroup's closing price and closing price adjusted for stock splits and dividend, as well as their corresponding returns series (computed by the concatenating method). Some notable features of these series are: (i) the price spikes just before 1998 (the year of the merger between Citicorp and Travelers Group) and exhibits some volatility before continuing to grow; (ii) there are two other major falls of the stock, corresponding to the 2000s Internet bubble bust and the Enron scandal (2002) as well as the recent GFC (just after the rescue package the firm received); (iii) the huge downward trends lead to the most volatile period in returns from 2008 to 2010. In addition, when comparing the closing and adjusted closing price series, a very strong impression is the extremely different scales on the two graphs. In subsequent analyses we shall focus on the 'adjusted' returns time series, in which the definition of the one-period continuously compounded returns r_t is given by:

$$r_t = \ln P_t - \ln P_{t-1}$$

where P_t is the adjusted closing price at time t.

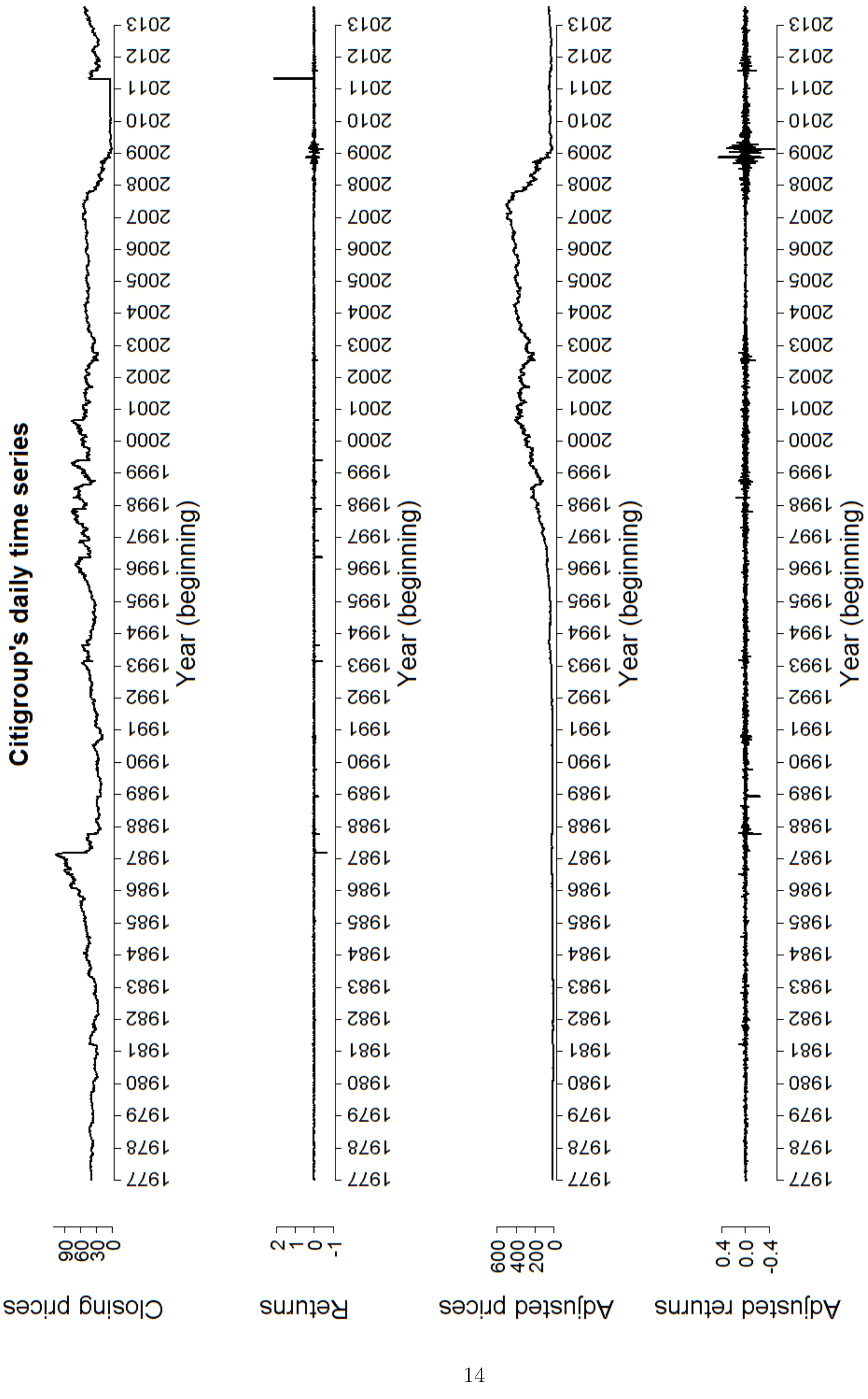


Figure 3: Time series plot of Citigroup daily data. From top to bottom: close price, returns (close-to-close), adjusted close price, returns (adjusted close-to-adjusted close). Data range from 03 Jan 1977 to 31 Jul 2013.

4.2 Leptokurtic returns distribution and outliers

A common feature of the distribution of financial returns (or residual returns) is its excessive kurtosis compared to a Gaussian distribution (whose kurtosis is 3). In other words the fourth central moment of these variables is usually greater than 3, implying a heavy-tailed distribution. To illustrate, we plot the histogram of the standardized returns $Z_t = \frac{r_t - \bar{r}}{s}$ where $\bar{r}$ and s are the sample mean and standard deviation of returns, respectively. When compared with other distributions such as the Gaussian, the Student-t (with degrees of freedom of 10) and the Generalized Errors Distribution (GED) we see that the distribution of our standardized returns exhibit excess kurtosis and slightly negative skewness, as shown in Figure 4.

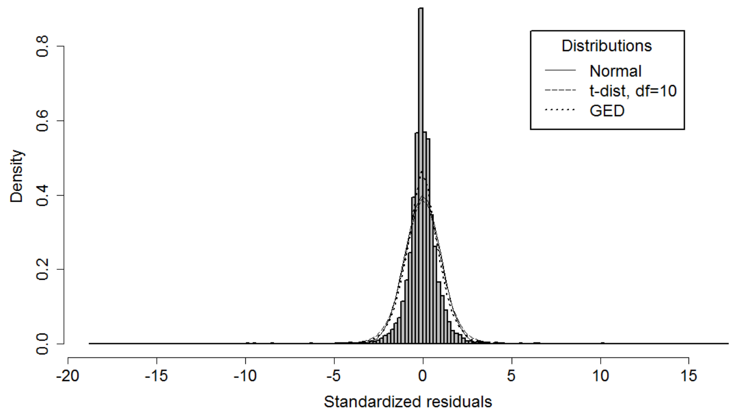


Figure 4: Histogram of standardized daily Citigroup returns for period 1977-2013, with lines indicating fitted normal, student and GED density functions superimposed.

We also note that since the fourth moment raises the variation to a power of four, it is very sensitive to extreme fluctuations of returns around the mean. Therefore, when examining kurtosis, it is desirable to winsorize our data, to make the empirical implications less susceptible to outliers. In particular, we truncate the extreme returns smaller than the 1% quantile and greater than the 99% quantile and consider these as outliers representing impacts of market crashes. Then we replace the negative extreme values by the 1% quantile and the positive extreme values by the 99% quantile. From Figure 5, when studying the boxplots of these 4 time series we can clearly tell the significant reduction of outliers when replacing the 'raw' data with winsorized data. Here the width of the boxes represents the Interquartile Range (IQR), or the difference between the upper quartile-UQ (or the 75% quantile) and lower quartile-LQ (or the 25% quantile). The lower and upper 'whiskers' indicate the values equal LQ $-$ 1.58 $\times$ IQR and UQ $+$ 1.58 $\times$ IQR, respectively (see e.g. (McGill et al. (1978), p.16) and (Chambers, 1983)). Any value falling outside of the range implied by these whiskers is considered an outlier. Also, when comparing original returns and winsorized returns there is not much difference between the position of the boxes or the median, as these are based on middle values and are robust to outliers. In addition, the mean of both original and winsorized returns (identified by the black lines in the boxes) are close to zero. Similar observations can be made with the volatility series, in addition to the pure positive outliers of the corresponding boxplots.

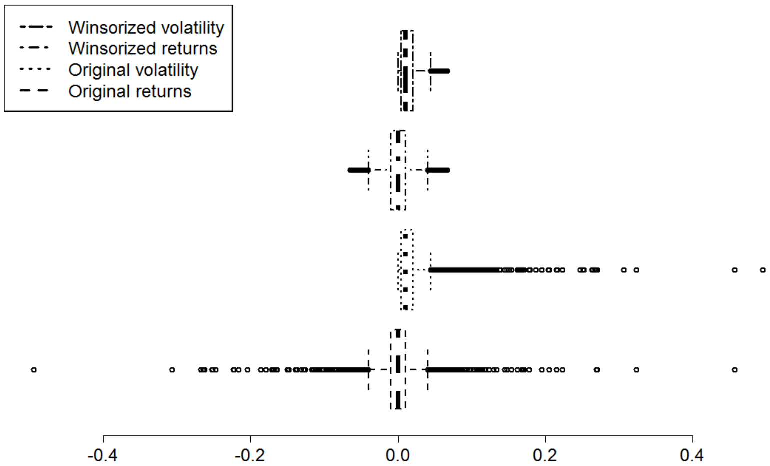


Figure 5: Boxplots of daily time series of Citigroup for period 1977-2013. The various dashed and dotted lines (in the middle of the boxes) indicate the respective median of each of the four series. Formulation of the boxes' range and whiskers can be found in (McGill et al., 1978).

The summary statistics of the raw returns (denoted as r_{raw}) and winsorized returns (denoted as r_{wins}) time series along with their corresponding volatility proxies (absolute returns) are reported in Table 2. The distribution of standardized winsorized returns, viz. r_{wins}, still exhibits robust excess kurtosis compared to a standard normal distribution. The Jarque-Bera test rejects the normality assumption for all 4 time series, implying a heavy-tailed distribution. In addition, the Ljung-Box test strongly suggests autocorrelation among these empirical returns as well as the corresponding volatilities (although the evidence is much weaker for returns). This will be explored in the next section. From this point on, we shall utilize the winsorized returns for further analyses. That is, unless stated otherwise, the daily data examined in subsequent studies are all based on the winsorized returns.

	r_{raw}	$\lvert r_{raw}\rvert$	r_{wins}	$\lvert r_{wins}\rvert$
Mean	0.000189	0.015805	0.0003052	0.014786
Median	0	0.010118	0	0.010118
Variance	0.000703	0.000453	0.0004308	0.000212
Skewness	-0.605176	6.576009	0.0246598	1.639483
Kurtosis	42.9454	84.3004	1.725833	2.58466
JB	710028.5(0.0000)	2800229(0.0000)	1147.528(0.0000)	6706.388(0.0000)
LB (21)	144.31(0.0000)	17018.05(0.0000)	52.3834(0.0000)	13161.6(0.0000)

Table 2: Summary statistics for Citigroup daily data for the period 03 Jan 1977 to 31 Jul 2013. p-values are reported in parentheses.

4.3 Long-run autocorrelation structure

When we examine Figure 3 the returns series appears to be mean stationary but not covariance stationary, i.e. while the mean does not deviate from zero, the variance of the process is itself time varying. This feature is reflected in the 'clusters' of volatility as illustrated by the absolute series. To better clarify this fact, we plot the autocorrelation function for the returns series together with its squared and absolute series in Figure 6.

For the returns series, although it is difficult to observe the overall significance of the ACF, the Ljung-Box portmanteau test in section 4.2 indicates there is serial correlation among returns up to lag 21, implying our empirical returns are not independent. For the volatility series, we can clearly see the hyperbolic decaying ACFs of the squared and absolute returns which retain their high significance for as far as lag 400.

Furthermore, when we increase the lag range, we observe a more clearly visible hyperbolic decaying pattern of the ACF functions of squared returns and absolute returns. This indicates a long-memory behaviour of these processes, and implies the long run impact of return shocks, which must be taken into account when modelling volatility.

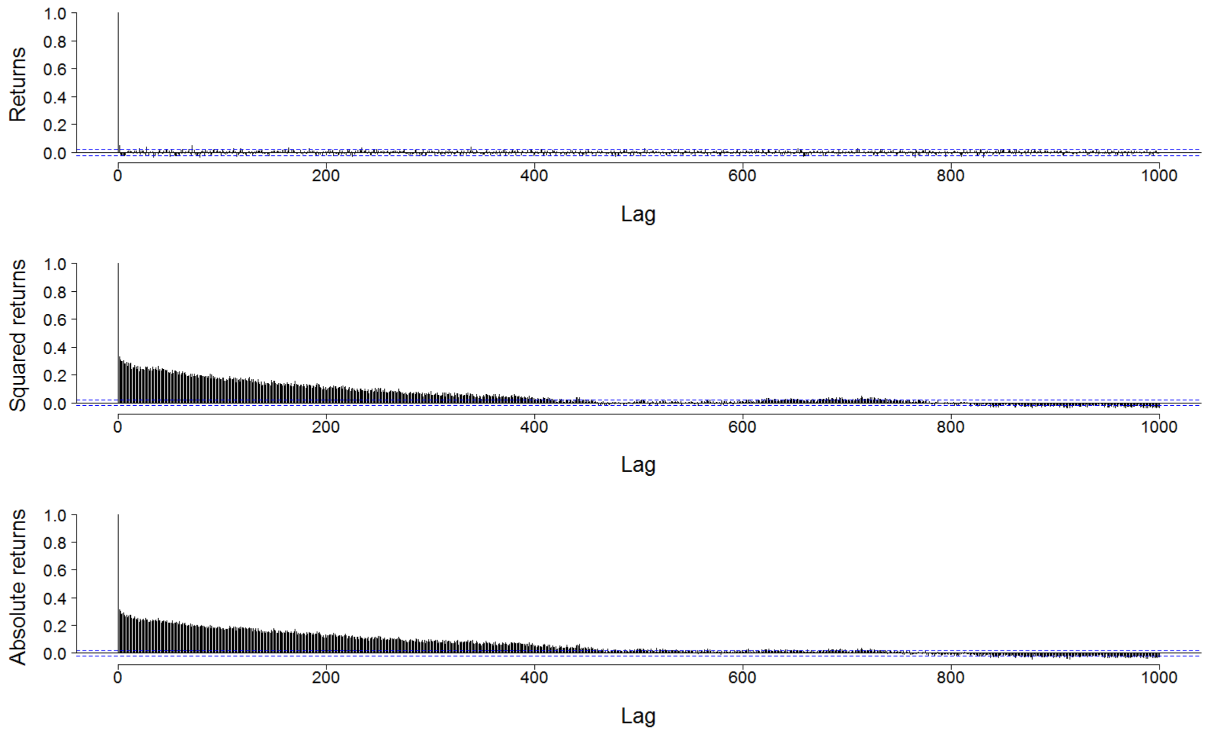


Figure 6: Correlograms of Citigroup daily time series for the period 1977-2013, up to 1000 lags. The dashed lines indicate the 95% confidence intervals.

4.4 Test for unit root non-stationarity

It is observed that unit root non-stationary time series (i.e. integrated processes) could exhibit slowly decaying ACFs similar to those of stationary long-memory processes. Therefore it might not be possible to distinguish the two type of processes using only the ACF (Brooks, 2002). Here we apply the Augmented Dickey-Fuller (DF) test for unit root stationary of our time series (see e.g. (Dickey and Fuller, 1979) and (Hamilton, 1994)). Specifically, we use the following model:

$$\Delta(X_t) = \alpha + \beta t + \gamma X_{t-1} + \sum_{i=1}^{p} \delta_i \Delta(X_{t-i}) + \epsilon_t$$

Here α is a constant and β is the coefficient of the time trend. Including both coefficients allow us to test for unit root with a drift and a deterministic time trend simultaneously. Unlike the original DF, the ADF adds the lagged differenced terms to account for the serial correlation up to order p in the data generating process which could invalidate the statistical inference of the DF test. The 'optimal' number of augmenting lags (p) is determined by minimizing the Akaike Information Criterion. The ADF test statistic is the t-statistic of the OLS estimate of γ. The null hypothesis of the ADF test is $\gamma = 0$. Intuitively, when $\gamma = 0$ is not rejected, the time series is not stationary, and the lagged level (X_{t-1}) cannot be used to predict the lagged change ($\Delta(X_t)$). As can be seen from table 3 the ADF test rejects the null hypothesis at any level of significance for all series. This means our returns, squared returns and absolute returns processes may be considered stationary, which, for our purposes, can be modelled by the long-memory DGP described in section 3.

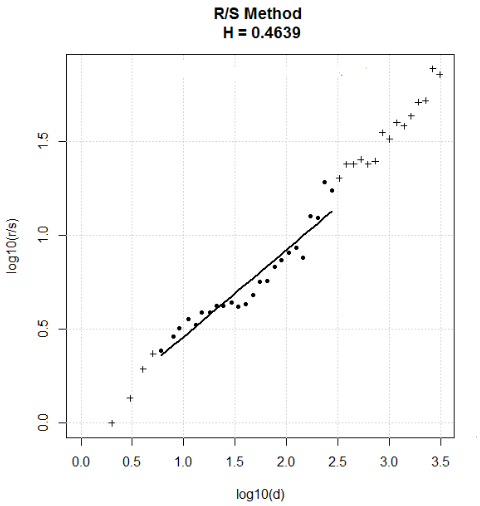
Null hypothesis $H_0 : \gamma = 0$			
Series	ADF stat	Critical value*	
Returns	-67.142	1%	-3.4593
Squared returns	-45.963	5%	-2.8738
Absolute returns	-46.95	10%	-2.5732

Table 3: Augmented Dickey-Fuller test for stationarity in daily time series.
(*) The critical values are obtained from (MacKinnon, 1994).

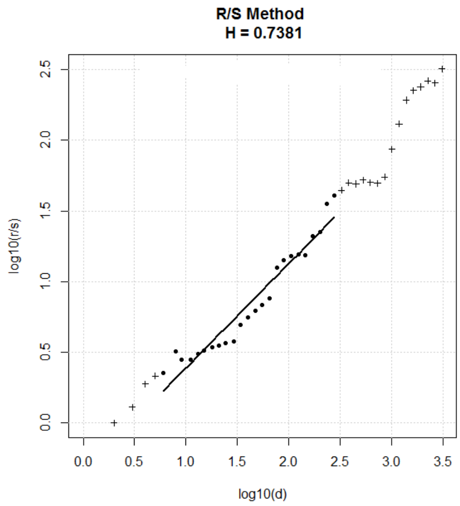
5 Empirical study

5.1 Estimating the Hurst index of Citigroup daily time series

Figure 7 reports the Hurst exponent estimated by the modified R/S method, for daily returns and volatility series of Citigroup. For the daily volatility process, overall we observe a fairly strong long run dependence: judging from the propositions introduced in subsection 3.2, our estimate of the Hurst index (0.7381) seems to indicate a certain degree of predictability of Citigroup's volatility. On the contrary, the returns series exhibits only weak dependence: the corresponding estimate of the Hurst index is in close proximity to 0.5, which is the theoretical value characterizing a random, uncorrelated process. For our purposes, it is sufficient to regard the returns series as a white noise process.



(a) Panel A: Returns



(b) Panel B: Volatility

Figure 7: Hurst index estimate of Citigroup's (winsorized) returns and volatility, using the (modified) R/S method.

The nature of the dependence structure of returns and volatility series is further supported by the visual features of their respective autocorrelation functions as shown in Figure 6. In particular, the absolute returns ACF decays at a hyperbolic rate whilst there is little/insignificant serial correlation among returns.

5.2 Time-varying Hurst index estimates

This subsection investigates the possibility of a time dependent long-memory parameter. From the previous section it can be concluded that the daily returns of Citigroup, Inc. follows a martingale and cannot be predicted based on past information, thus supporting the general assumption of a (weakly) efficient market. However, this argument may only be valid from an aggregated perspective. As (Mitra, 2012) pointed out, it is not unusual to observe deviation from market efficiency in the form of a Hurst exponent being different from 0.5 at a local scale (See also (Whitcher and Jensen, 2000)). This deviation is manifested in short-term trending and/or mean-reverting behaviour which could be capitalized on by technical analysts from time to time. Along the same lines, (Qian and Rasheed, 2004) find that the daily returns series of the Dow-Jones Industrial Average index (DJIA) exhibits remarkable difference in its dependence structure over time: strong persistence during the period 1930-1980 and mean-reversion during the period 1980-2004.

5.2.1 Pre-determining the confidence interval of Hurst index estimate

Following previous discussions, in general we would expect $H = 0.5$ to be the 'efficient market benchmark' when examining stationary time series with a large sample size. However, because we plan to apply our estimator over a relatively short time window (much shorter than the full sample studied earlier), we need a re-definition of the point estimate and corresponding confidence interval for the Hurst index associated with an efficient market. One straightforward way to do this is to perform a simplified Monte Carlo simulation with the following steps: (i) generate 100,000 random time series, each of which has length $n = 1,024$ (we use the fractional Gaussian noise DGP with $H = 0.5$ as a representation of a 'true' random process); (ii) calculate the sample average and standard deviation. These values will help us specify the new benchmark point estimate and the corresponding confidence interval. The idea is to approximate the expected value of a random variable (Hurst index) by the mean of a large number of random draws of that variable.

Based on these calculations, our new benchmark for an efficient market shall be $H = 0.5608$ and the 95% confidence interval (under Gaussian distribution assumption) for this estimate is $0.5608 \pm 1.96 \times 0.0392$. That is, we can be 95% confident in stating that the true population value of the Hurst index of a length $n = 1024$ random time series lies within the interval $[0.4839, 0.6378]$.

Consequently, when studying our time-varying estimate series, we classify the behaviour of the series into four categories: strong persistence when $0.64 < H < 1$; weak persistence when $0.56 < H < 0.64$; weak anti-persistence when $0.48 < H < 0.56$ and strong anti-persistence when $0 < H < 0.48$. These bounds are similar to those used by (Mitra, 2012) and (Hull and McGroarty, 2013).

5.2.2 Computing rolling Hurst index estimates

Next, to examine the time-varying nature of Hurst exponent estimate we follow (Qian and Rasheed, 2004)'s approach, i.e. we adopt a 'rolling sample' estimate of Hurst index. First, the estimate is computed for the first time window of 1024 days (approximately 4 years) using the Rescaled range method. According to (Qian and Rasheed, 2007), one of the reasons for the choice of a length of $n = 1024$ days is that (Peters, 1994) had documented a four-year cycle in the DJIA index. Although we only examine Citigroup stock price here, this cycle could be relevant since Citigroup was one of DJIA components until recently. Then, the window is rolled one-period ahead and we re-estimate the Hurst index for the next day. This gives us a time-varying daily series of Hurst estimates with length equals to that of the input returns series minus the first 1024 observations, for a total of 8204 observations starting from the beginning of 1981. As can be seen from Figure 8, the estimates of Hurst index vary widely from 0.4571 to 0.6550. However, statistically speaking we cannot say that the true value of H is different from 0.56, which reconfirms the market efficiency implied from an aggregated perspective. For most of the 8204 observations we observe weak anti-persistence behaviour (58.31%) and weak persistence behaviour (40.22%) (although there is some evidence of strong anti-persistence (mean reversion) in mid 1998 and strong persistence at some times from 2008 to 2010).

In addition, there are two notable sharp changes in the dependence structure of the returns series, namely: (i) the quick fall from and rise back to $H = 0.56$ observed from mid 1997 to end of 1999; (ii) the steep rise from 0.48 to 0.64 observed from mid 2007 to mid 2008. Interestingly, these periods are associated with the two of the largest crises in the financial world. Relating these events to the corresponding movements of the stock price (see Figure 3 from subsection

4.1) we can see that the values of the Hurst index estimate can indeed reflect the persistence of returns: for example, the price level from the period 2007-2008 was all but plummeting while that of the period 1997-1999 experienced a steady rise followed by a large, short-lived drop (possibly corresponding to the major stock split in June 1998, around the time of Citigroup's merger with Travellers) before bouncing back to the upward trend. Likewise we observe large clusters in volatility around the height of the GFC while the same cannot be said about the earlier period.

Another possible explanation for the strong mean reverting behaviour observed in 1997-1999 would be some 'spill-over' effect from the Asian financial crisis. A large number of studies considered the quick recovery of the Asian markets back to the mean level (in 1999) to be a result of a temporary over-reaction which was quickly corrected (see e.g. (Patel and Sarkar, 1998), (Malliaropulos and Priestley, 1999) and (Fujii, 2002)). In this regard, the crisis in 2008 is inherently different, as it reflects a fundamental weakness of the U.S. financial system, rather than a short-term over-reaction, thus explaining the sharp increase in Hurst index from 2007 when the market's confidence continued to fall.

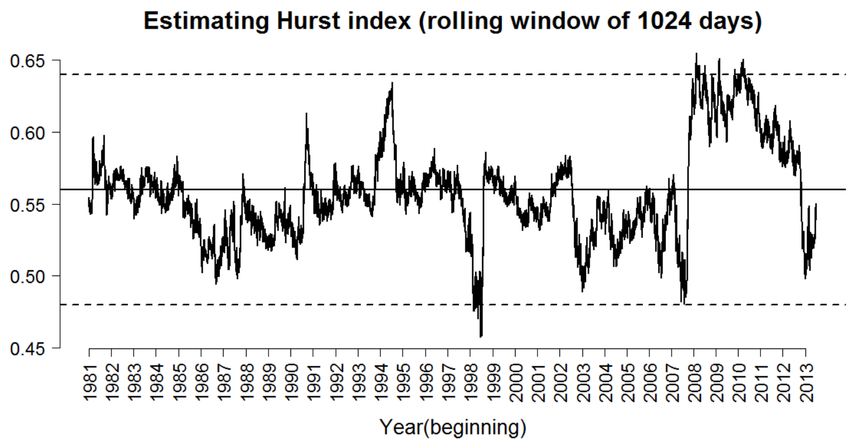


Figure 8: Time-varying Hurst index estimate of daily returns of Citigroup, Inc. Using the R/S method, the Hurst index is computed for a rolling window of 1024 days, starting from 03 Jan 1977. The first observation is computed at the beginning of 1981. The line indicates the efficient market benchmark for the Hurst index estimate ($H = 0.56$), whilst the dashed lines approximate the 'upper' and 'lower' bounds of the 95% confidence interval ($H = 0.48$ and $H = 0.64$, respectively).

5.2.3 Comparing Hurst index estimates with returns from different trading strategies

In the following discussion we shall look at how to reconcile the story of the long-run dependence exhibited throughout this paper with some of the most commonly used trading strategies that take advantage of such dependence structures. Specifically we shall examine the smoothing techniques related to the Simple Moving Average (SMA) (or alternatively, the Exponentially Weighted Moving Average (EWMA)). In its simplest form the (equally weighted) SMAs are computed by taking the average of the most recent sequence of closing prices over a specific number of days; then this 'window' is rolled forward one period to compute the next observation, hence the name 'moving average'. Therefore all moving averages are lagged indicators as they are computed from past data. As such, an n-day SMA series is defined as:

$$p_t^{(n)} = \frac{1}{n} \sum_{i=t-(n-1)}^{t} p_i$$

Without any doubt these averaging metrics are among the oldest and most popularly used indicators of existing trending pattern of stock prices. It is widely documented that many stock traders rely, to some extent, on the signal produced by the simple moving average to build their strategies and make investment decisions. This procedure constitutes a broader family of "trading rules". (Brock et al., 1992) argue that "*[...]Technical analysis is considered by many to be the original form of investment analysis, dating back to the 1800s. [...]These techniques for discovering hidden relations in stock returns can range from extremely simple to quite elaborate.*" (p.1731).

For our purposes, we follow a simple trading rule: whenever the closing stock price moves above an average level (for example a 10-day SMA) then it indicates a buy signal. A sell signal appears as soon as stock price comes below the moving average value. In his discussion, (Taylor, 2005) added another type of signal: neutral, which is triggered when the difference between short and long SMAs falls within a certain bandwidth, thus this difference is not enough to give a precise view about the trend. In our simple case this bandwidth is set to zero. We shall keep buying/selling the stock until the signal turns to a 'sell'/'buy' and vice versa. To simplify the investigation, we assume that there is no trading cost involved, which is not so impractical when studying heavily traded stocks. Because the SMA smooths the original price series to reveal a trend, these strategies are commonly associated with terms such as 'trend-following' or 'trend-identity' (Reuters Limited, 1999). Alternative strategies can be employed in which a mid term average (e.g. 50-day) crossing a long term average (e.g. 200-day): a buy (sell) signal is triggered when the short term average crosses above (below) the long term one. Such a strategy is known as a 'cross-over'. Another version of it is formed with multiple long term SMAs crossing a shorter-term SMA, which was referred to as a 'ribbon cross-over'. The signals given by this type of technical analysis are robust in the sense that they are strong indicators of an upward/downward trend, as can be seen in Figure 9, in which a sell signal triggered in August 2007 was followed by a steep downward trend of actual price at the beginning of the GFC. Traders can look toward such signals to determine the entry and exit points according to their preference. In addition, one of the main reasons why this basic, yet powerful tool has become very popular among short-term traders is its objectivity (as the signals are not governed by investors' subjective evaluations). Apparently, although the formation of the signals is objective, the choice of which indicator to use is entirely a matter of preference. Specifically, the speed of adjustment to price changes and/or the number of signals generated depend heavily on the time window of the moving average(s) chosen.

In practice this method uses closing prices as inputs to compute trading profits; however, to have comparable implications with our returns and volatility series we choose to starting with adjusted closing prices instead. To illustrate, we plot four particular SMA series (which are commonly used in practice) along side Citigroup's daily adjusted closing price for the year 2007 in Figure 9. As can be seen, the obvious weakness of SMAs is that they are lagged time series, i.e. their upward/downward movements are always lagged compare to the original price series, the lag effect increasing with the length of the time window on which the SMA is based. In general, a shorter-term average system 'moves' faster and create more signals, though the reliability of said signals may not be as good as those created by longer-term moving averages, in terms of identifying a genuine trend. As a result, traders opting for a longer-term system tend to have more room to 'surf the waves' whilst a shorter term average may produce too many 'false signals' and may prevent optimal profit earnings due to premature buying/selling. In any case, long-term indicators are more effective when the trending behaviour is strong.

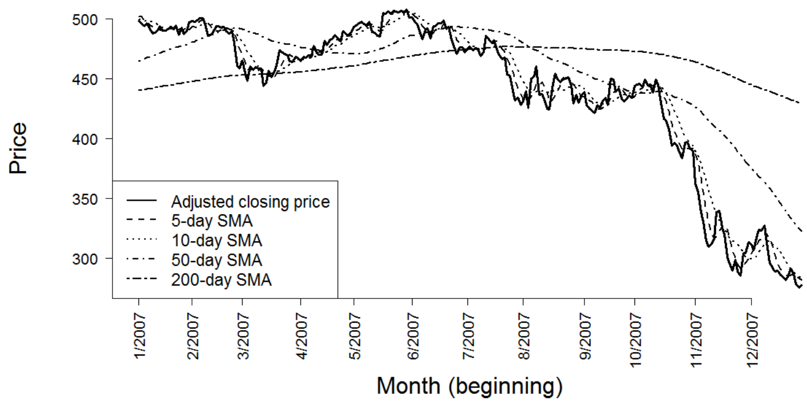


Figure 9: Plot of adjusted closing price for the year 2007, together with 5-day, 10-day, 50-day and 200-day SMA. To maintain the number of observations the SMAs are computed from a sample longer than the year 2007, since all SMAs are lagged series.

(Mitra, 2012) proposed a simple procedure to illustrate the relevance of Hurst exponent estimates in conjunction with the performance trading strategies based on these averaged indicators. The procedure contains three steps:

- Step 1: First, they split the daily sample into non-overlapping sub-periods of 60 days, then estimate the Hurst index for each sub-period. We find this approach unfavourable because the small size of the sub-periods may hamper the accuracy of Hurst index estimates. Therefore, we modified this step by incorporating (Qian and Rasheed, 2004)'s approach. That is, our time-varying daily Hurst index estimates are based on a rolling window of 1024 days. We denote this series as H_{roll}.

- Step 2: Next, we follow the simple trading rule discussed above. In particular, at the beginning of any day we compare the closing adjusted price with the SMA computed from prices of n days preceding that day (n=10, 20, 50 or 200). Then we would sell/buy whenever the price is lower/higher than the SMA and compute the corresponding returns series based on the winsorized adjusted returns (r_{wins}) constructed in subsection 4.2. For example, if the original return is -0.02 on a specific day then at the end of the day we will realize a return of 0.02 (or −0.02) should we receive a sell (or buy) signal at the beginning of that day (based on information from the previous day's closing price). We repeat this step with the four SMAs representing short-term (5 days and 10 days), mid term (50 days) and long term (200 days) strategies. We denote these series as SMA5dret, SMA10dret, SMA50dret and SMA200dret, respectively.

- Step 3: The final step is to assess the cross-correlation between the Hurst estimates from step 1 and the corresponding returns obtained from each of the trend-following strategies described in step 2. We also need to exclude the first 1024 observations in each returns series when computing the correlation coefficients. Naturally the correlation should give us an idea of how well our long range dependence parameter reflects the trending behaviour of stock prices captured by strategies designed to capitalize on such behaviour.

In Table 4 we report the cross-correlation coefficients between H_{roll} and the winsorized daily returns (r_{wins}) and the four SMAs, in addition to corresponding statistics for the test of no significant correlation. In agreement with the findings of (Mitra, 2012), the positive linear relationship clearly implies that the Hurst index can capture the trending characteristics of a financial time series. As we can see, the null hypothesis of no correlation can reasonably be rejected (at the 1% level) in three cases: SMA5dret, SMA10dret and SMA50dret, all of which exhibit a positive correlation coefficient. Because r_{wins} is not constructed from a trend-following rule, there is no evidence of significant correlation between this original returns series and H_{roll}. Interestingly, the longer the time window the SMA is based on, the weaker the linear relation, and the 200-day SMA returns show no correlation with H_{roll}. Overall we can conclude that the short term SMAs tend to provide more informative signals than longer term ones.

Additionally, we compute the average trading returns corresponding to the four intervals of Hurst index estimate mentioned earlier, i.e. ($H < 0.48$, $0.48 < H < 0.56$, $0.56 < H < 0.64$ and $H > 0.64$). As can be seen in Table 5, in general higher H values are associated with higher trading returns. Intuitively, the trend-following strategies work best when the trends are strong. In addition, the average returns obtained from such strategies are all higher than the average winsorized returns. All in all, this reconfirms the indicative value of the Hurst index.

When analysing the time-varying nature of the Hurst index series and its correlation with trading returns, we can spot a clear linkage to the most important implications of the "Adaptive Market Hypothesis" proposed by (Lo, 2004). Specifically, although persistently profitable trading strategies are impractical, exploitable opportunities do exist from time to time, and certain strategies tend to succeed in certain environments. Though we have not explored further the profound underlying determinants of said environments (i.e. what really makes a strategy profitable, and why?), our evidence supports the view of a dynamic, evolutionary perspective of market efficiency, rather than the "*inexorable trend towards higher efficiency predicted by the EMH*" (Lo, 2004).

In any case, the SMAs used in this study, as well as the rolling window chosen (1024 days) are merely illustrative and should be taken as reference only. It is possible that with other specifications we should obtain different interpretations. The implications from trading techniques based on other indicators (such as the EWMA) could also be important. In addition, the question of whether mean-reverting trading strategies perform better when the series exhibit mean-reverting behaviour remains open. Therefore, for possible future research, a more complete survey is desirable, to provide a better understanding of these intriguing observations.

The main finding in this subsection is that financial returns exhibit some predictability which may be revealed by its long term dependence structure. Adopting a different, but related approach, (Brock et al., 1992) and (Taylor, 2005) examine the distributional properties of buy and sell returns (obtained from similar trading rules) and concluded that historical prices are informative about future returns as long as these distributions are different. Intuitively, the so-called trading rules dictate that investors hold more stocks when the recent expected returns (proxied by short-term SMAs) are higher than past expected returns (proxied by long-term SMAs). Therefore, "*there is some predictability in the returns process whenever these expectations are fulfilled.*" ((Taylor, 2005), p.159).

CCF of H_{roll} and	(a) r_{wins}	(b) SMA5dret	(c) SMA10dret	(d) SMA50dret	(e) SMA200dret
Coefficient	-0.0201	0.1436**	0.0847**	0.0337**	0.0093
t-stat	-1.8226	13.1484	7.7019	3.0550	0.8447
p-value	0.0684	0.0000	0.0000	0.0023	0.3983

Table 4: Cross-correlation coefficients between the rolling Hurst index estimate series and different returns series, accompanied by corresponding statistical inference metrics. The null hypothesis is that the true (population) correlation coefficient is zero.
(**) indicates significance at the 1% level.

	Number of observations	Average trading returns				
		r_{wins}	SMA5dret	SMA10dret	SMA50dret	SMA200dret
$H < 0.48$	38	0.00304	0.01511	0.01297	0.00597	0.00304
$0.48 < H < 0.56$	4784	0.00062	0.00896	0.00659	0.00307	0.00151
$0.56 < H < 0.64$	3300	-0.00012	0.01339	0.00925	0.00454	0.00221
$H > 0.64$	82	0.00459	0.02366	0.01632	0.00425	-0.0016
Total	**8204**	**0.00037**	**0.01092**	**0.00779**	**0.00369**	**0.00177**

Table 5: Average trading returns corresponding to different values of the Hurst index estimates.

5.3 A contrasting example with simulated data

From the previous study we can see that the cross-correlation between Hurst index estimates and trading returns are relatively weak, albeit statistically significant. To emphasize the importance of this finding, we consider similar correlation by replacing actual data with stock prices simulated by a geometric Brownian motion (gBm). To do this, we first use the actual data to estimate the parameters for this gBm process, then simulate a whole new sample with initial price equal to the first price observation in the actual sample. We expect to observe no significant correlation between the Hurst index estimated from the simulated returns series and the trading returns constructed from strategies applied to simulated price.

To begin, we assume daily stock price S_t follows a gBm:

$$dS_t = \mu\ S_t dt + \sigma\ S_t dW_t$$

where μ_t and σ_t are the drift and volatility, respectively. W_t is a Wiener process or standard Brownian motion. Itô's lemma shows that when stock price follows a gBm we have:

$$\Delta\,\mathrm{log}S_t = \mathrm{log}S_{t+s} - \mathrm{log}S_t \sim N\left(\left[\mu - \frac{1}{2}\sigma^2\right]s, \sigma^2 s\right) \tag{7}$$

Now we fit the following linear regression model to the (discrete time) actual data:

$$\mathrm{log}S_{t_{n+1}} - \mathrm{log}S_{t_n} = \alpha + \epsilon_{n+1} \quad \text{with} \quad \epsilon_{n+1} \sim N(0, \phi^2) \tag{8}$$

so that we have

$$\mathrm{log}S_{t_{n+1}} - \mathrm{log}S_{t_n} \sim N(\alpha, \phi^2) \tag{9}$$

From expressions 7 and 9 we have the estimates of α and ϕ^2 should satisfy:

$$\hat{\alpha} = \left(\hat{\mu} - \frac{1}{2}\hat{\sigma}^2\right)\ \Delta t \quad \text{and} \quad \hat{\phi}^2 = \hat{\sigma}^2\Delta t \tag{10}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the needed estimates of μ and σ. Δt is the discrete time interval, which is set to $1/252$, representing a working day.

Utilizing the settings discussed above, we proceed to simulate the gBm by the following four steps:

- From the actual (adjusted closing) price series, construct the variable $X = \Delta\mathrm{log}S_{t_{n+1}}$
- Obtain the estimates of model 8 as:

$$\hat{\alpha} = E(X) = E(\alpha + \epsilon_{n+1}) = 0.00019 \quad ; \quad \hat{\phi}^2 = \mathrm{Var}(X) = \mathrm{Var}(\epsilon_{n+1}) = 0.000703$$

- Compute the values:

$$\hat{\mu} = \frac{2\hat{\alpha} + \hat{\phi}^2}{2\Delta t} = 0.1365 \quad ; \quad \hat{\sigma} = \frac{\hat{\phi}}{\sqrt{\Delta t}} = 0.4210$$

- Simulate the gBm using the Euler approximation in discrete time:

$$\mathrm{log}S_{t_{n+1}} = \mathrm{log}S_{t_n} + \hat{\mu}\Delta t + \hat{\sigma} Z_{n+1}\sqrt{\Delta t} \quad (n = 0, 1, 2, \dots)$$

  where Z_{n+1} are random numbers drawn from the standard normal distribution. Finally, we have a price series generated by the following model:

$$S_{t_{n+1}} = S_{t_n}\ \exp\ (\hat{\mu}\Delta t + \hat{\sigma} Z_{n+1}\sqrt{\Delta t}) = S_{t_n}\ \exp\ (0.1365 \times 1/252 + 0.4210 \times Z_{n+1}\sqrt{1/252})$$

By fixing a simulated price series as input, then reconstructing the winsorized returns series and repeating the procedures described in subsection 5.2.3, we find insignificant (and mostly negative) cross-correlations between the Hurst estimates and the trading returns obtained from different strategies. Results are reported in Table 6. This clearly indicates that the significant correlation we found in subsection 5.2.3, albeit weakly, is not spurious. Intuitively, modelling the returns process by the gBm may not be appropriate when the actual data exhibit (local) long-range dependence and/or predictability.

CCF of H_{roll} and	(a) r_{wins}	(b) SMA5dret	(c) SMA10dret	(d) SMA50dret	(e) SMA200dret
Coefficient	-0.0057	0.0048	-0.0013	-0.0007	-0.0019
t-stat	-0.5186	0.4363	-0.1240	-0.0722	-0.1801
p-value	0.6041	0.6626	0.9013	0.9424	0.8571

Table 6: For simulated stock prices following a gBm: cross-correlation coefficients between the rolling Hurst index estimate series and different returns series, accompanied by corresponding statistical inference metrics. The null hypothesis is that the true (population) correlation coefficient is zero.

6 Concluding remarks

Our primary contribution in this paper revolves around the estimation of the long-memory parameter of financial time series: in line with previous researchers (e.g. (Peters, 1996)), in general we find a significant long-run dependency in volatility, with the Hurst index estimated to be in the vicinity of 0.7; while the returns series does not exhibit such behaviour, with a Hurst index estimate indistinguishable from 0.5. This reconfirms the market efficiency at an aggregate level, as returns follow a martingale and are generally independent, whereas volatility exhibits a certain degree of predictability (which also results in the clustering patterns observed).

However, when we examine our data from a local perspective, that is, via a rolling window of 4 years, we see that the long-run dependence behaviour of returns is not time-invariant. In fact, the daily Hurst index estimated from a rolling window of 4 years varies widely, being higher whenever returns' dependence increases. This is consistent with recent findings which emphasize the connection between the long-run dependence nature of financial time series, the trending behaviour of stock prices and the evolution of market efficiency ((Lo, 2004), (Mitra, 2012), (Hull and McGroarty, 2013)).

Furthermore, the time-varying Hurst index series is positively correlated with the returns series obtained from trading strategies based on the relative position of the simple moving averages compared to actual stock price. As these strategies are designed to detect and capitalise on the short term 'trends' in the market, our result clearly implies that the Hurst index is a good indicator of such trending behaviour and may be of interest to stock traders seeking to improve their strategies. For example, traders may opt for trend-following rules (mean-reverting rules) with stocks exhibiting Hurst index higher (lower) than 0.5. All things considered, from a statistical viewpoint, these results might shed new light on the credibility of technical analysis.

As a final precaution, our study, for all intents and purposes, is designed to explore the volatility structure of only one company, Citigroup Inc., which, despite having a heavily traded stock reflecting important market fluctuations, may possess unique capital structure properties that affect our findings and make them biased. In addition, for many reasons the firm was highly impacted by the GFC, more than any other financial service company. This means our results may have limited general validity. On the other hand, the fact remains that it is the special position of Citigroup in the global finance system that provides us with a worthy candidate to study the inter-relationship between returns and volatility. To extend our paper's implications to a wider group, in the future, we could conduct event studies to focus on some particular periods that can be related to the company's financial structure, or we could study a group of companies that share Citigroup's intrinsic characteristics. Additionally, we could organize similar investigations on various indexes, from both developed and emerging markets, to draw more general conclusions.

References

Andersen, T. G. and Bollerslev, T. (1997). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long run in high frequency returns. *Journal of Finance*, 52(3):975–1005.

Beran, J. (1994). *Statistics for long-memory processes*, volume 61. Chapman & Hall, New York.

Bollerslev, T. and Wright, J. H. (2000). Semiparametric estimation of long-memory volatility dependencies: the role of high-frequency data. *Journal of Econometrics*, 98(1):81–106.

Brock, W., Lakonishok, J., and LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance*, 47(5):pp. 1731–1764.

Brooks, C. (2002). *Introductory Econometrics for Finance.* Cambridge University Press, Cambridge, U.K.

Carmona, P. and Coutin, L. (1998). Fractional Brownian motion and the Markov property. *Electronic Communications in Probability*, 3.

Cavalcante, J. and Assaf, A. (2004). Long-range dependence in the returns and volatility of the Brazilian stock market. *European Review of Economics and Finance*, 3(4):5–22.

Chambers, J. (1983). *Graphical methods for data analysis.* Chapman & Hall statistics series. Wadsworth International Group.

Cont, R. (2005). Long range dependence in financial markets. In *Fractals in Engineering*, pages 159–179. Springer.

Cox, D. R. (1984). *Long-range Dependence: a Review.* Statistics: an appraisal, (H.A. David and H.T. David, eds.). Iowa State University Press.

Dickey, D. A. and Fuller, W. A. (1979). Distributions of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431.

Dieker, A. B. and Mandjes, M. (2003). On spectral simulation of fractional Brownian motion. *Probability in the Engineering and Informational Sciences*, 17(3):417–434.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106.

Engle, R. F. and Patton, A. (2001). What good is a volatility model? *Quantitative Finance*, 1(2):237–245.

Fama, E. F. (1965). The Behavior of stock market prices. *The Journal of Business*, 38(1):34–105.

Fujii, E. (2002). Exchange rate and price adjustments in the aftermath of the Asian crisis. *International Journal of Finance & Economics*, 7(1):1–14.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3):393–408.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press, Princeton, N.J.

Heyde, C. C. (2002). On modes of long-range dependence. *Journal of Applied Probability*, 39(4):882–888.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.

Hull, M. and McGroarty, F. (2013). Do emerging markets become more efficient as they develop? Long-memory persistence in equity indices. *Emerging Markets Review*, 18:45–61.

Hurst, H. (1951). Long term storage capacity of reservoirs. *Transaction of the American Society of Civil Engineer*, 116:770–799.

Lo, A. W. (1991). Long-term memory in stock market prices. *Econometrica*, 59(5):1279–1313.

Lo, A. W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management*, 30(5):15–29.

Lo, A. W. and MacKinlay, C. A. (1999). *A Non-random Walk Down Wall Street.* Princeton University Press, Princeton, NJ.

MacKinnon, J. G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business & Economic Statistics*, 12(2):167–176.

Malliaropulos, D. and Priestley, R. (1999). Mean reversion in Southeast Asian stock markets. *Journal of Empirical Finance*, 6(4):355–384.

Mandelbrot, B. (1966). Forecasts of future prices, unbiased markets, and martingale models. *Journal of Business*, 39(S1):242–255.

Mandelbrot, B. (1972). Statistical methodology for nonperiodic cycles: from the covariance to R/S analysis. In *Annals of Economic and Social Measurement*, volume 1 of *NBER Chapters*, pages 259–290. National Bureau of Economic Research, Inc.

Mandelbrot, B. and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10:422–437.

Mandelbrot, B. and Wallis, J. R. (1968). Noah, Joseph, and operational hydrology. *Water Resources Research*, 4(5):909–918.

Mandelbrot, B. (1971). A fast fractional Gaussian noise generator. *Water Resources Research*, 7(3):543–553.

McGill, R., Tukey, J. W., and Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, 32(1):12–16.

Mitra, S. K. (2012). Is Hurst exponent value useful in forecasting financial time series? *Asian Social Science*, 8(8):111–120.

Newey, W. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–08.

Patel, S. A. and Sarkar, A. (1998). Crises in developed and emerging stock markets. *Financial Analysts Journal*, 54(6):50–61.

Peters, E. (1994). *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics.* Wiley Finance. John Wiley & Sons, New York.

Peters, E. (1996). *Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility.* John Wiley & Sons, New York.

Qian, B. and Rasheed, K. (2004). Hurst exponent and financial market predictability. In *IASTED conference on Financial Engineering and Applications (FEA 2004)*, pages 203–209.

Qian, B. and Rasheed, K. (2007). Stock market prediction with multiple classifiers. *Applied Intelligence*, 26(1):25–33.

Reuters Limited (1999). *An Introduction to Technical Analysis.* John Wiley & Sons, New York.

Samuelson, P. A. (1965). Proof that Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, 6:41–49.

Taqqu, M., Teverovsky, V., and Willinger, W. (1995). Estimators for long-range dependence: An empirical study. *Fractals - an Interdisciplinary Journal on The Complex Geometry of Nature*, 3(4):785–798.

Taylor, S. (2005). *Asset Price Dynamics, Volatility, and Prediction.* Princeton University Press, Princeton, N.J.

Teverovsky, V., Taqqu, M. S., and Willinger, W. (1999). A critical look at Lo's modified R/ S statistic. *Journal of Statistical Planning and Inference*, 80(1):211–227.

Whitcher, B. and Jensen, M. J. (2000). Wavelet estimation of a local long memory parameter. *Exploration Geophysics*, 31(2):94.

Wuertz, D. (2013). *fArma: ARMA Time Series Modelling.* R package version 3.0.1.

Xu, S.-j. and Jin, X.-j. (2009). *Predicting drastic drop in Chinese stock market with local Hurst exponent.* Moscow, Russia.