DEPENDENCE STRUCTURE IN FINANCIAL TIME SERIES: APPLICATIONS AND EVIDENCE FROM WAVELET ANALYSIS

by

Long Hai Vo * †

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Master of Commerce
in Finance

Victoria University of Wellington
2014

*Student at School of Economics and Finance, Faculty of Commerce, Victoria University of Wellington. ID number: 300259065.
Tel.: +84 1266 727 527 (Vietnam)
E-mail: longvoqnu@gmail.com

† *Acknowledgement:* I sincerely thank my supervisor, Dr. Leigh Roberts, for his enormous encouragement and patience during the course of my thesis. I am also grateful to Prof. Graeme Guthrie and Dr. Toby Daglish for their insightful and constructive comments, especially on the empirical findings. All remaining errors are my own.

Abstract

Conventional time series theory and spectral analysis have independently achieved significant popularity in mainstream economics and finance research over long periods. However, the fact remains that each is somewhat lacking if the other is absent. To overcome this problem, a new methodology, wavelet analysis, has been developed to capture all the information localized in time and in frequency, which provides us with an ideal tool to study non-stationary time series. This paper aims to explore the application of a variety of wavelet-based methodologies in conjunction with conventional techniques, such as the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models and long-memory parameter estimates, in analysing the short and long term dependence structure of financial returns and volatility. Specifically, by studying the long-memory property of these time series we hope to identify the source of their possible predictability. Above all else, we document the indispensable role of trading activities associated with low frequencies in determining the long-run dependence of volatility. It follows that GARCH models incorporating long-memory and asymmetric returns-volatility dynamics can provide reasonably accurate volatility forecasts. Additionally, the persistence parameter of returns, represented by the Hurst index, is observed to be correlated to trading profits obtained from typical technical rules designed to detect and capitalize on existing trending behaviour of stock prices. This implies that the Hurst index can be used as a good indicator of the long-memory characteristic of the market, which in turn drives such trending behaviour.

This thesis is dedicated to my parents, Don Vo and Hoa Nguyen, who never give up hope on me; and to my dearest sister, Linh Vo, who taught me that even the most daunting task can be accomplished if it is done one step at a time.

Contents

List of Figures

Page

Page

List of Tables

Page

Chapter 1

Introduction

1.1 Theoretical motivation

The relationship between risk and return is the mechanism underpinning every single economic transaction that takes place in any trading environment. The financial market is no exception. Investors and traders constantly revise their decisions and strategies on the basis of varying estimates of risk and return. Traditionally, in a financial market context, the relatively broad concept of risk is illustrated by the notion of "volatility" which describes the variability of possible outcomes of an investment. Since the birth of modern financial markets, many authors have attempted to explained the complex interrelation between return and volatility and its evolution through time. This has resulted in an impressive body of literature on financial time series.

Empirical findings from these researches generally reveal that every set of financial returns exhibit a number of specific characteristics aptly named "stylized facts". In essence, these features are related to the statistical properties of the return variable, and reflect the mechanisms driving the return generating process. It is important to keep in mind that the number of stylized facts changes with new discoveries in the literature, and there exists variability from market to market. Arguably, the most prominent among these "facts" are the non-normal, heavy-tailed distribution of returns; the insignificant correlation between consecutive returns (apart from very short-run correlation as a result of microstructure effects); and the significant positive autocorrelation between their absolute values which is very persistent in the long run. In particular, this behaviour of volatility was observed to cause persistence in the dependence structure of the absolute returns and squared returns (a fact generally referred to as "long-memory"). From the traditional approach, by setting the volatility of returns to be time-varying, many econometric specifications have achieved a satisfactory fit to the empirical data. In this paper we shall primarily focus on the interpretation of this type of dependence structure while examining various related stylized facts.

Searching for a model that could appropriately account for these stylized facts has been a daunting task that has fascinated generations of academics in

the field. From a time series perspective, the two prominent features implied by the volatility of financial returns are what were known as clustering and long-memory. These phenomena have been documented in a wide range of financial disciplines. The phrase 'clustering' is used to describe the situation when large/small price fluctuations are most likely followed by a movement similar in scale. Early documentations concerning clusters of movements of the price process include Mandelbrot (1963) and Fama (1965), whilst similar observations are extensively reported by Baillie, Bollerslev, and Mikkelsen (1996), Schwert (1989) and Chou (1988). 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. All in all, these well-established stylized facts generally imply a certain degree of predictability of financial volatility, as opposed to randomly generated returns which are uncorrelated under the efficient market assumptions. To provide a complete account for all stylized facts associated with the volatility of finance returns is well beyond the scope of this paper. As stated, the persistent dependence structure will be our primary focus. Nevertheless, other related facts including mean-reversion (generally interpreted as the existence of a long-run level of volatility) and asymmetric impact of positive/negative innovations shall be addressed in later discussions.

The most widely acclaimed method for modelling volatility is a framework designed to capture the varying nature of the returns' conditional variance, known as the Autoregressive Conditional Heteroskedasticity (ARCH). The idea was proposed by Engle (1982) and Bollerslev (1986). The original framework has been intensively expanded into a family of models that contributed greatly to the business of modelling volatility [1]. In general, these models can capture the evolution of the volatility process at a single time horizon. We can colloquially say that by adopting such approach, we are able to "horizontally" fit a time series with respect to the stylized features of its volatility. As such, both volatility clustering and long-memory can be classified as "*horizontal dependence*" behaviours in the time domain (Gençay, Gradojevic, Selçuk, and Whitcher (2010)).

However, this particular approach to volatility has an inherent weakness:

[1] See, for example, Engle, Lilien, and Robins (1987), Nelson (1991) and Glosten, Jagannathan, and Runkle (1993), among many others, for some of the 'iconic' GARCH specifications.

it is limited in the sense that it specifies only an individual trading frequency, within a single time scale. Simply speaking, each financial time series is specified with a trading horizon chosen from the perspective of the traders (be it daily, monthly or annually etc.). The problem with this approach is that we do not know how volatilities evolves across different time horizons [2]. As Gençay et al. (2010) pointed out, a model capable of describing the data generating process at the daily frequency may not do so well explaining hourly movements of the data. Acknowledging this weakness gives rise to some questions: how do we determine if there exists a relationship between the variability of intraday returns and that of returns recorded at lower frequencies? How, and why, do the volatility estimates vary across multiple time horizons/scales?

More importantly, what new insight would this type of "vertical" analysis provide us regarding the primary forces driving the volatility process, and to add a complement to the well-observed "horizontal" facts? We speak of "vertical" in the sense of looking at multiple frequencies simultaneously (while still preserving information of the time domain) via an advanced mathematical tool called Multi-resolution Analysis (MRA) that shall be the workhorse of our study in the frequency domain. Utilizing this technique, an important observation was made by Gençay et al. (2010): low volatility recorded at long time horizons is more likely to be followed by low volatility at shorter horizons, but this is not necessarily true for the high regimes of volatility. These authors consider this as a new stylized fact of volatility and termed it "*asymmetric vertical dependence*". Their findings not only reconfirm the link between low-frequency and high-frequency representations of volatility documented by Muller et al. (1997), Zumbach and Lynch (2001) and Borland, Bouchaud, Muzy, and Zumbach (2005), but also clarify this link by examining the asymmetric nature of the information flow: impact of low frequency on high frequency trading activities is stronger when associated with low volatility rather than high volatility regimes.

As mentioned above, it is observed that trading horizons vary with traders' investment strategies and thus correspond to different trading frequencies (Dacorogna, Gençay, Muller, Olsen, and Pictet (2001)). For example, low frequency transactions can be viewed as associated to daily or monthly horizons. In contrast, intraday traders implement orders during daytime with very frequent transactions, in some cases (e.g. for market makers) between very small time intervals such as minutes, seconds or 'continuously' between logical units of information known as "ticks". Unlike long-term traders, high-frequency market participants

[2] In the relevant literature, as well as in the current paper, the following pairs of terminologies are interchangeably adopted: high/low frequencies, short/long time horizons and small/large scales (or fine/coarse scales).

may need to constantly adjust their positions for various purposes including hedging or finding optimal asset allocation. For such a heterogeneous, multi-frequency trading environment, long-term traders generally are only concerned about large price movements (which are normally observed over long time intervals). In contrast, short-term traders need to pay attention not only to those large movements but also to high-frequency, highly active market fluctuations over short intervals. In other words, information conveyed by volatility 'shocks' flows from long-term to short-term horizons and not the other way around. This also implies that trading activities at certain low frequencies are expected to be the most influential components of the aggregated volatility process (although when measuring variance, most variance can be associated with high frequencies).

Additionally, the hidden Markov model used by Gençay et al. (2010) indicates frequent regime switching (between high and low volatility states/regimes) at high frequencies, which implies the market 'calms down' earlier than when being viewed at lower frequencies. Intuitively, if there is a large movement in low-frequency activities, its is likely to be reflected in the volatility process at much higher frequencies. On the other hand, a high volatility at long horizons does not necessarily lead to a high volatility at shorter horizons. Alternatively, the empirical findings of Gençay et al. (2010) can be summed up in two points: for a volatility process, in general (i) the duration of regimes is longer at lower frequencies and (ii) the duration of low regimes is longer than that of high regimes at all frequencies.

With regard to the discussions above, we learn that comprehensive modelling of volatility necessitates an approach which can incorporate information of volatility's behaviour in both frequency and time domains. Observed financial returns are not homogeneous in the sense that they can be broken down into various time series associated with different horizons or time scales (or resolutions - a terminology that is more common in engineering studies). As a result, examining intraday financial data at different scales will naturally provide important insights regarding the fundamental components of volatility dynamics. To this end, a mathematical tool, the MRA, has been developed based on the wavelet theory. Naturally, in order to conduct analysis in the pre-described multi-scale context, acquiring multi-scaled data is a crucial first step. In the next subsection we explore the role of high-frequency data in our research.

1.1.1 The importance of intraday data on revealing volatility dependence structure

In the most recent decades, the availability of financial data has played an increasingly indispensable role in facilitating the development of quantitative models aiming to construct volatility theories. This point of view can be underscored by simply considering the parallel evolution of data frequencies and the corresponding diversity of such models. Linear model families such as the Autoregressive Integrated Moving Average (ARIMA) had been used to model monthly or yearly data from the 1970s. When daily and weekly data became available in the 1980s, discoveries of varying conditional variance of returns called for the non-linear branch of ARCH models. In time, even the conventional GARCH did little in terms of explaining the new range of properties exhibited by the more recently available high-frequency data (one such property is the 'intraday seasonality' which will be further explored in the later chapters).

As its name suggested, high-frequency financial data are recorded more frequently than the 'standard' data (e.g. daily data). Generally, the data used in published studies of financial time series analysis are both low-frequency and regularly spaced, while availability of high-frequency data has not resulted in a comprehensive literature until recently. Not surprisingly, the low sampling rate of data can help reduce the costs of collection and storage significantly, and the equally spaced data points made it simpler to developed statistical inferences, though with limited confidence (Dacorogna et al. (2001)).

Unfortunately, most high-frequency primary data are based on information arriving randomly and directly from markets, such as quoted prices from transactions. This feature makes it very difficult to study these data using methods that are designed for analysing equally spaced time series. On the other hand, one critical advantage introduced by huge influxes of high-frequency [3] information, in terms of goodness-of-fit tests, is that we can readily reduce the number of testing specifications. This is because statistical uncertainty decreases as the frequency of observations rises (Muller et al. (1995)). Based on this argument, a heterogeneous ARCH (or HARCH) model was developed by Muller et al. (1997) to better reflect the impact of various market participants on variance estimates.

1.1.1.1 Improvement on volatility estimates

Unlike low-frequency data, at the base level the intradaily information is updated much more frequently depending on the strategies of traders, with the

[3] Hereafter high-frequency refers to any frequency higher than daily, i.e. intraday data.

disruption periods being as small as fractions of a second. Naturally, when being observed, such high-frequency data provide us with a unique opportunity to study financial markets at different time scales, from the finest to the coarsest. Improvement of volatility estimates, as documented by the following researchers, is one among many advantages supporting this approach.

Merton (1980) argued that high frequency sampling significantly improves the accuracy of the time-varying variance estimate of the price process, although this paper was restricted to only monthly and daily data. In line with this argument, Martens (2001) explicitly modelled intraday volatility and concluded that the forecast of daily volatility using intraday returns is superior to that of the traditional approach using daily returns. In addition, the major contribution of the seminal paper by Nelson (1992) is that the accuracy of volatility estimates can be greatly improved with sufficiently frequently sampled data, even with very short calendar time spans. The evidence of similar insights is intuitively provided by Andersen and Bollerslev (1998), in which the authors give a clear explanation of the poor out-of-sample predictive power of the GARCH model family despite adequate in-sample fit. The problem, as they pointed out, is not from the well-specified models themselves, but the ex-post volatility measurements. The conventionally used squared return innovations was proved to be a very noisy, albeit unbiased, proxy for the latent volatility. This motivates these authors to introduce the so-called realized volatility which is defied as cumulative squared intraday returns. They also noted that "this measure converges to genuine measurement of latent volatility factor" (effectively making volatility "observable") given that the sampling frequency increases indefinitely. This important notion is extensively discussed in the excellent review of Poon and Granger (2003). Evidently, using high-frequency data enables us to compute superior estimates of lower-frequency volatilities.

Many studies came to the common conclusion that due to specific market characteristics, realized volatility measured from returns computed at very small intervals (in particular less than five minutes) is more likely to be contaminated by microstructure effects (most notably non-synchronous trading and bid-ask bounce) (Poon and Granger (2003)). Andersen, Bollerslev, and Cai (2000) examined the intraday volatility of Japanese stock market, and documented a distinctive U-shaped pattern closely following intraday trading periods, in which volatility tends to be high in the opening and closing of market while decreasing throughout the day. This article shed light on the role of microstructures such as the impact of asymmetric information on price formation.

For our purpose, although it is infeasible and not necessarily advantageous

to acquire data at extremely high frequencies because of data limitations and the presence of microstructure effects, the intraday data is expected to provide a superior estimate of volatility as well as a better understanding of the long-run dependence behaviour of volatility (some analyses with the same goal are Bollerslev and Wright (2000), Andersen and Bollerslev (1997a) and Andersen, Bollerslev, Diebold, and Labys (1999)).

1.1.1.2 Revelation of volatility dependence structure

Another important aspect of intraday data is that they can be used to investigate the difference in the dependence structure of volatility at different frequencies. In some cases, such as in a GARCH(1,1) model, inter-temporal aggregation is theoretically proved to be able to preserve volatility structure (See Drost and Nijman (1993)). However, empirical evidence generally suggests otherwise: volatility modelled at high frequencies exhibits much stronger persistence than at low frequencies (Diebold (1988)). This notion is, in principle, in line with Nelson (1991) and is supported by Glosten et al. (1993), whose main finding is that persistence in volatility seems to diminish significantly when moving from daily to monthly data. Likewise, Andersen and Bollerslev (1997a) argued that the long-memory behaviour observed in daily data also characterizes the volatility process at high frequencies, after filtering out all intraday dependencies. Therefore, for volatility time series, increasing observation frequency will improve the estimate of long-memory parameter much more effectively than increasing the number of time points.

With regard to the examination of high-frequency stylized facts, Andersen and Bollerslev (1997a) noted the sharp contrast between the highly dependent dynamic in daily volatility to a less profound long-term dependence among intraday volatility. According to these authors, the seemingly contradictory evidence of volatility persistence in daily and intradaily data is attributable to the fact that intraday volatility can be regarded as "*[The] aggregation of numerous constituent component processes; some with very short-run decay rates [which dominate in intraday returns] and others possessing much longer-run dependencies*" (p.1002). In turn, the multiple volatility components at the intradaily frequencies are a result of numerous heterogeneous information arrival processes. From a different angle, Jensen and Whitcher (2000) capture the presence of both short and long-memory (as a result of unexpected as well as anticipated news) with a non-stationary, but locally stationary, stochastic volatility estimator. Their primary finding is a time-varying long-memory parameter which exhibits an intradaily pattern: being highest and lowest in accordance with the opening and closing

of foreign exchanges. These authors open up the debate of whether this strong time-of-the-day effect is caused by public macroeconomic news announced at the beginning of some markets, or the capitalization of informed traders at the end of other markets. In any case, the link between volatility dependence structure and the pattern in which new information arrives is central in understanding market operating dynamics. Furthermore, the evidence that long-memory parameter peaks at the close of most active trading exchanges documented by Jensen and Whitcher (2000) and heterogeneous volatility structures captured by Andersen and Bollerslev (1997a) both point to the association between the persistence of volatility and less active, less volatile trading sessions. It would be interesting, then, to examine this relationship across different time horizons. Assuming the most volatile behaviour is observed at the highest frequencies, we expect to obtain smaller persistence measurements at these frequencies.

1.1.2 From volatility to volume and information flows

A theory aiming to provide intuition behind the long-run dependence structure of volatility is the Mixture of Distributions Hypothesis (MDH) proposed by Clark (1973). The theory has enjoyed increasing popularity since the 1970s because it reconciles two related stories: long-run dynamics of volatility and strong contemporanerous correlation between volatility and volume. Like the former, the latter fact is also documented in numerous papers. The principle of MDH is that both the volatility and trading volume processes are subordinate to an unobserved variable known as information flow. Tauchen and Pitts (1983), among others, defined information flow as the rate at which new information arrives. Bollerslev and Jubinski (1999) point out that price changes are governed by this underlying latent variable: good/bad news invokes price increase/decrease, both of which result in higher-than-average trading volume and volatility. Therefore the number of price changes varies during the day depending on the number of news items arriving that day.

Early tests of the MDH yield mixed interpretations: some studies support it (Harris (1986)), some argue that short-run joint dynamics of volume and volatility cannot be explained by a single latent information-arrival process (Richardson and Smith (1994)), others attribute the validity of MDH to the choice of distributional assumptions (Andersen (1996)). After taking previous studies into account, the main argument of Bollerslev and Jubinski (1999) is that the MDH should be examined from a long-run perspective rather than focusing on the 'misspecified' short-run dynamics. Their supportive evidence is the very similar degree of fractional difference of three processes: information, volatility and volume. Fleming

and Kirby (2011) revisit this proposition using more precise volatility measures and find that while the evidence of similar degree of long memory is much weaker, the strong correlation between volume and volatility is confirmed. As a result volume can be used to estimate volatility when high-frequency data are unavailable.

1.1.3 Stylized facts of high-frequency financial returns

The most notable fact of financial returns is perhaps the Non-normal (or Non-Gaussianity) or the heavy tailed distribution. Earlier empirical analyses dated as far back as the 1970s had repeatedly placed much emphasis on documenting related visual features and numerous implications of returns' distribution such as its approximate asymmetry, the heavy tails and its high peak (see, for example, Mandelbrot (1963) and Fama (1965)).

Dacorogna et al. (2001) and Taylor (2005) documented and summarized several stylized facts of volatility for both low-frequency and high-frequency returns. Because there are variations in their orders and wordings, we believe it is helpful to list the propositions in group of 'traditional' categories, namely: facts about distributions, autocorrelations, seasonality and scaling properties:

- The middle price is subject to microstructure effects at the highest frequency.
- Intraday returns exhibit distribution with heavy tails, with increasing kurtosis proportional to increase in observation frequency.
- Intraday returns from traded assets are almost uncorrelated, a part from a significant first-order correlation
- There exists seasonal volatility clusters distinctive to different days of the week.
- There exists short bursts of volatility following arrival of major macroeconomic news.

Our main interest is placed on the long-memory property exhibited by measures of daily and intradaily volatility. Poon and Granger (2003) assert that slow decay autocorrelation of variances is a very well-established stylized fact throughout the literature (refer to Granger, Spear, and Ding (2000) for further reviews). According to Andersen and Bollerslev (1997a), long-term dependency is an inherent characteristic of the underlying volatility generating process. These authors also attribute the inconsistent conclusions about the level of volatility dependence to the fact that volatility process is actually a sum of multiple components.

Volatility estimates are therefore affected by either short-run components (when using high-frequency data) or long-run ones (when using daily data). However, with appropriate restrictions imposed, the aggregated volatility process always exhibits long memory. Christensen and Nielsen (2007) and Fleming and Kirby (2011) as well as many academics having experience in the field agreed upon the feasibility of describing realized volatility with a fractionally integrated process where the long memory parameter d is within the range of 0.3 - 0.5. However, they also documented short-lived modest impacts of volatility changes on stock prices. In any case, volatility long-run dependence structure possesses a complex nature.

Why is the existence of long range dependence crucial to studies of financial time series? One problem with the classical asset pricing theories is pointed out by Lo (1991): whenever such dependence structure exists among financial returns, the traditional statistical inference for tests of the capital asset pricing model and the arbitrage pricing theory, both of which rely on a martingale returns process, is no longer valid. Furthermore, the larger the long-memory effect, the greater the impact of shocks on future volatility. In practice, however, financial returns are almost always observed to be uncorrelated and thus exhibit very little long-run dependence.

1.1.4 The rise of Wavelet analysis

We acknowledge that the well-known family of ARCH and its related extensions are often successful in fitting the daily data as well as capturing the major stylized facts of volatility, but as Vuorenmaa (2005) argued, fixing a specific time horizon for which to analyse stock market volatility seems to be inadequate. According to the heterogeneous market hypothesis proposed by Muller et al. (1995), the market is "fractal" in the sense that it is composed of different trading participants. Strategies of these market participants are based on past information associated with their respective time horizons, from seconds (in the case of market makers), minutes (for speculators) to years (e.g. investment activities of pension funds, insurance agencies and possibly some types of hedged funds). Although we could attribute this heterogeneous nature to several factors such as different market perception, level of risk aversion and financial constraints, or other technical factors (e.g. geographic locations), sensitivity to time-scales could be representative of most of these differences, as shown by Muller et al. (1995). This hypothesis of multi-layered investment horizons clarifies the existence of the heterogeneity in the stock market, and gave rise to the heterogeneous ARCH model (Muller et al. (1997)). Recently this hypothesis has been validated both theoretically and

empirically and played an important role in delivering a better understanding of the stylized facts, in particular the highly persistent behaviour of volatility, at a level never observed before. With the growing popularity of high-frequency data disclosure, the urge to examine and bring this new insight into modelling volatility became both a possible and standard approach. That is, we must be able to analyse our data at multiple time scales, or frequencies, if we seek to capture the true volatility dynamics.

The frequency analysis is illustrated via studies of the frequency spectrum, which enables us to identify the time-scales aspect of a stationary series. Standard non-parametric spectral analysis transforms the autocorrelation function into a power spectrum. Through comparison with the parametric GARCH models, this transformation reveals the various "global" sinusoid components of the stationary signal. In other words the objective of frequency analysis is to decompose the original stationary signal into representative underlying components. However, the same empirical problem with standard GARCH specification persists: because 'real' financial data are generally not stationary, observing global sinusoids is most likely not feasible. Instead of focusing on global "large wave" sinusoids, an alternative usage of "small and varying waves", which will be elaborated on in later chapters, was proposed to better capture the local behaviour of empirical time series. The need for a flexible, non-parametric multi-scale analytical tool to study the non-stationary nature of the heterogeneous market is the main motivation for the application of wavelet theory.

This paper primarily revolves around a relatively novel algebraic framework known as wavelet-based multi-resolution analysis (MRA hereafter). According to Gençay et al. (2010), *"[A] successful method to describe the market dynamics at different time scales (monthly, daily, hourly, etc.) must be able to separate each time-scale component from the observed data"* (p.897). The wavelet analysis is a potential candidate to fit into this picture, making it possible to break the fluctuations in signals down into components associated with different frequencies, or scales. Some of the most recent notable achievements regarding the application of wavelet in volatility studies in particular and asset pricing in general, as well as in revision of volatility theories, can be found in the following list. For our purposes, from these papers we take note of a consensus centred in a frequency-dependent volatility-return (risk-reward) interrelation, with a multi-scale perspective:

- In and Kim (2005) investigate the Fisher hypothesis regarding positive relationship between stock returns and inflation. Their evidence supports this hypothesis only at the lowest and highest scale length (corresponding to the shortest and longest horizon) while rejecting it at intermediate scales.

This helps explain the contradictory results of previous researches which independently studied data at different frequencies.

- In, Kim, and Gençay (2011) examine the dynamic of the optimal risk exposure of stocks over different investment horizons with wavelet methodology. Specifically, they find that for lower and moderate levels of risk aversion, optimal asset allocation shifts away from growth stocks in favour of value stocks when the horizon increases. This result implies that value stocks are less risky than growth stocks at long horizons.

- Lee, Kim, and Lee (2011) applied wavelet analysis on high frequency data to review two hypotheses of volatility dynamic: the leverage effect and feedback effect. They found the return-volatility relationship at the smallest scales is different from that at larger scales and there exists a long-memory behaviour of daily realized volatility.

- Gherman, Terebes, and Borda (2012) combined the strength of a wavelet transform and GJR-GARCH model to yield improved volatility forecasts. In particular the outputs of their wavelet transform are used as inputs for GJR-GARCH. Subsequently they perform an inverse transform to get prediction of returns and prices.

- Barunik and Vacha (2012b) proposed an advanced wavelet-based estimator that provides significant improvement in forecasting volatility via a realized jump GARCH framework. They documented improved volatility forecasts with the use of time-frequency volatility measurements. They also found that most volatility is associated with higher frequencies rather than lower ones. The theoretical motivation for their estimator can be found in Barunik and Vacha (2012a).

In recent years wavelet theory has provided the foundation of more and more breakthroughs in physical sciences and engineering, and to some extent in the medical science. Despite its very extensive application in these fields, this methodology has recognized astoundingly little popularity in the mainstream empirical finance researches. Exceptions to the rule are Vuorenmaa (2005), who proposed a multi-scale approach to studying the change in returns' variance and Whitcher and Jensen (2000) who examine a wavelet-based estimator of a time-varying long-memory parameter in the context of locally stationary financial volatility process. A promising indicator of changes to come may be the recent publication of In and Kim (2013), which, as its predecessor written by Gençay, Selçuk, and Whitcher

(2002) did, nicely introduces the power of wavelet analysis to many traditional problems in economics and finance in the form of interesting and novel approaches.

1.2 Research structural description

To reach our goal of improved understanding of the long run behaviour of financial volatility, we utilize different quantitative modelling techniques, which offer complementary perspectives. This section briefly describes the overall approach we intend to take in our study. Throughout most of the empirical analyses, daily and intradaily data are studied in a 'parallel' manner, both of which shall be the object of our various experiments.

First, we adopt a family of GARCH models to examine and analyse the prominent stylized facts of volatility of financial returns. In particular, models that are specifically designed to capture asymmetric impacts of share price changes on volatility are emphasized. Second, we employ several well-established estimators capable of detecting the long-run dependence behaviour of volatility. In addition, we provide analyses regarding how the findings of these two methodologies strengthen each other. Finally, we extend our study by using the same approaches to the multi-scale representations of the original data. Specifically we use wavelet based decomposition technique to compare volatility patterns across different time horizons. This is to identify and locate any changes in the overall volatility structure during the span of data sample. Furthermore, we need to clarify the 'causal' relationship between changes observed in low frequency (long horizon) trading and high frequency (short horizon) trading. Our speculation is that the variation at low frequencies are more likely to determine that at the high frequencies while the converse might not be true.

We choose to study in detail just one specific company in the financial service industry, viz., Citigroup Inc. This is of course rather limiting, in that much of the literature is concerned with stock indices, especially the Standard & Poors 500; and the papers analysing individual companies did not necessarily choose Citigroup. Nevertheless, given that we wished to go beyond the volatility of stock indices in this thesis, Citigroup was a logical choice. On the one hand, many of the innovations introduced to the world of banking and finance over the last two centuries either sprang directly from Citigroup's initiatives, or were taken on board by Citigroup after being started by other banks. Since it was the largest bank in the world during the twentieth century, and well into the 21^{st}, Citigroup has had an enormous influence on the finance sector in general and banking in particular. Its impact more recently has also been strong, albeit for the wrong reasons, since it played a major role in the meltdown of the finance sector during

the 2008 Global Financial Crisis. These facts justify our argument that the factors determining the complex returns-volatility dynamics of the finance sector, and that of the market, could be reasonably revealed by examining Citigroup data. Further details about Citigroup are given in Appendix D.

With these objectives in mind, the rest of the paper assumes the following structure: Chapter 2 discusses various parametric techniques aiming to estimate the long-run dependence parameter, which is followed by the overview of the GARCH frameworks in Chapter 3. These two chapters serve as theoretical background for later empirical studies. In Chapter 4 we first describe the summary statistics of our daily and intraday data as well as important data treatment adjusting for effects that could otherwise distort the implications of our results, such as returns outliers, microstructure effect and intraday seasonality. The primary focus of Chapter 4 shall be the examination of the dynamics of the volatility data generating process across time and scale simultaneously; chief among which are the asymmetric, negative relationship between returns and volatility and the time-varying long-memory property of returns. Subsequently we show how the estimation of this time-varying long-memory help reveals predictability in price changes, which can be captured by conventional tools of technical analysis. Chapter 5 offers some concluding remarks and comments on the general implications of our study, as well as opening up some debates for future research.

The main empirical results reported in Chapter 4 are obtained via the application of wavelet methodology in conjunction with conventional techniques such as GARCH models and long-memory estimates (some of which are discussed in Appendix B). To avoid dilution by technical details, a general discussion of the theory of multi-scale analysis and its relevance to our paper is presented in Appendix A rather than in the main content. For our purpose, an ideal wavelet basis for studying financial time series should be sufficiently long, symmetric and smooth. It is observed that a major portion of existing literature features the LA(8) MODWT wavelet which handles the boundary conditions using the 'periodic' method. This wavelet has an acceptable balance of the above mentioned factors. It is therefore our choice unless explicitly stated otherwise. The implementation of various wavelet methodologies described is made primarily via the recently introduced R package waveslim (Whitcher (2012)).

Chapter 2

Long-memory estimating methods

Preamble

In the previous chapter we mentioned the well-known phenomenon of long term dependence structure for volatility, which is attributable to a long-memory data generating process. Evidence of this fact is often illustrated by the slow hyperbolic decay rate of the autocorrelation function of empirical time series in the physical sciences. To be specific, this type of process is defined with 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/Hurst index, and is named after the hydrologist H. E. Hurst who first analysed the presence and measurement of long-memory behaviour in stochastic processes. In the following Section 2.1, we shall cover the fundamentals of typical stationary long-memory processes, as well as their interconnections, to provide a theoretical background for studying our volatility process in later chapters. Section 2.2 then demonstrates a plethora of well-established and robust methodologies aimed to detect and estimate the degree of long-memory characterized by the level of the Hurst exponent H. Finally, in Section 2.3, by separately applying each of these methods on simulated data, we are able to confirm that the wavelet-based maximum likelihood estimator, which is formulated by Gençay et al. (2002), generally outperforms other methods in most of the simulated experiments, and is arguably inferior to none.

2.1 Introduction to stationary fractional difference processes

2.1.1 Time series and stationarity

Conventional analysis of stochastic processes uses a sequence of random variables ordered by a time index (denoted as $\{X_t\}$ $(t = 0, \pm 1, \pm 2, \dots)$) as a representation of the processes that generate a realized data set $\{x_1, x_2, \dots, x_n\}$. This time-ordered set of observations is called a time series. Here we are interested in

the time series represented by a sequence of financial returns $\{r_t\}$. If the joint distribution of $\{r_{t_1}, r_{t_2}, ..., r_{t_k}\}$ is identical to that of $\{r_{t_1+\tau}, r_{t_2+\tau}, ..., r_{t_k+\tau}\}$ for all $\tau \geq 1$ then we have a strictly stationary time series. Generally this strong assumption is violated in practice, and was replaced with a weaker version, which requires (i) $E(r_t) = \mu$ and (ii) $\text{Cov}(r_t, r_{t-l}) = \gamma_l$ (with l an integer). In other words, $\{r_t\}$ is said to be weakly stationary when its mean is a constant and its covariance is time-invariant, depending only on the distance between two time points (l) but not on the time point itself (t). Because of its connection to the second central moment, many authors referred to this category as second-order stationarity or covariance stationarity.

The simplest case of a stationary process is a white noise, which is a zero mean, uncorrelated process (meaning the autocorrelations among random variables constituting the process are all zero). A special case of white noise processes is specified with stronger assumptions: independent and identical normally distributed variables with zero mean and constant variance or i.i.d $\text{N}(0, \sigma^2)$ (Tsay (2001)). This process is then called a Gaussian white noise. A continuous stochastic process for which increments are Gaussian white noise is known as a Brownian motion (Bm) or a standard Wiener process (denoted as B_t), which is considered "the building block" of modern mathematical financial theories. That is, the change of process B_t over an infinitesimal time interval $[t, t + dt)$ is defined as $dB_t \sim$ i.i.d $N(0, dt)$.

2.1.2 Definition of long memory

A definition of a broad long-memory class can be found in Beran (1994), which states that such process is defined if its autocorrelation function $\rho(l) = \dfrac{\text{Cov}(X_i, X_{i+l})}{\text{Var}(X_i)}$ is not 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 (in general we have $\alpha = 2 - 2H$). In this case the decay is said to represent a "power law". Additionally, larger H implies stronger long-range dependence, or more persistent impact of past events on present events. Conventional statisti-

cal inferences of processes exhibiting this feature can be dramatically altered [4]. The long-memory processes are contrasted with the short-memory class, which exhibits summable and exponential decaying covariances (which was also termed short-range dependence or weak dependence). Lo (1991) made a clear distinction between these two classes, asserting that the short-memory behaviour is characterized by the fact that "*[...] the maximal dependence between events at any two dates becomes trivially small as the time span between those two dates increases.*" (p.1281). In other words, the rate at which dependence decays is very high for processes exhibiting short-run dependence.

To maintain broad consistency with previous surveys, in our paper the term 'long-memory' is interchangeably used with 'long-run dependence', 'long-term dependence' and 'long range dependence'. In other words, we note that the differences between these terms are negligible.

2.1.3 Some popular stationary long memory processes

Mandelbrot and his colleagues (see e.g. Mandelbrot and Wallis (1968) and Mandelbrot and Van Ness (1968)) were among the first to propose 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 is then viewed as a 'middle ground', i.e. I(d) where $0 < d < 1$. We show later that the fractional difference parameter d is related to Hurst exponent by the simple equality $d = H - 0.5$. The simplest way to distinguish the three types of processes is to look at the 'typical shape' of their respective autocorrelation functions: infinite persistence (for I(0)), exponential decay (for I(1)) and hyperbolic decay (for I(d)). In the following discussions we shall show that the autocorrelation of an I(d) process, though having various representations, assumes the general form of $\rho(l) = Cl^{-\alpha}$ for some constant C and integer α which is proportional to d and H.

[4] 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 inferences to be meaningful. However, with long-range dependence introduced by a slowly decaying $\rho(l)$, the corresponding 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.

2.1.3.1 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 $E[B_H^2(t)] = t^{2H}$ and covariance:

$$\gamma(s,t) = E[B_H(s)B_H(t)] = \frac{1}{2}(s^{2H} + t^{2H} - |s-t|^{2H}) \tag{2.1.1}$$

Carmona and Coutin (1998) provided a brief introduction to the Fractional Brownian process, which was termed a centred 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.

Similar to standard Brownian motion, fBm has several important properties:

- It has stationary increments, meaning the distribution of the difference between the realizations of this process is time-invariant, or $dB_H(t) = [B_H(t+dt) - B_H(t)] \sim B_H(dt)$.
- It is self-similar, which means $\{B_H(at)\}$ has the same distribution as $\{|a|^H B_H(t)\}$. We can colloquially say the process 'repeats' itself when comparing it with a scaled representation of itself (a is a scaling factor). The index $H \in [0,1]$ is called the self-similarity parameter. We already noted this parameter is also labelled as the "Hurst exponent" when the process exhibits long-range dependence.

Unlike the Bm, the fBm allows the increments to be correlated. With regards to the relationship between fBm and the standard Bm, we noted that when $H = 0.5$ then the fBm reverts to the Bm of which the data are 'randomly' generated, and there is no long-range dependence in the process's dynamics. Meanwhile, $H < 0.5$ means the increments of $\{B_H\}$ are negatively correlated (corresponding to the case of an anti-persistent process). Conversely, we have a positively correlated process if $H > 0.5$. This dependence structure implies clustering behaviour of the process. This property is crucial for modelling volatility: increases/decreases in past steps suggest increases/decreases in subsequent steps. Analogously, for $H > 0.5$, we have a long-memory pattern, as the covariance between the increments located far away from each other asymptotically decreases to zero at a slow rate:

$$\text{Cov}\,[B_H(t) - B_H(0), B_H(t+s) - B_H(s)] \sim s^{2H-2} \text{ when } s \to \infty \tag{2.1.2}$$

2.1.3.2 Fractional Gaussian noise (fGn)

As our direct application is for discrete data, it is only logical to move from differencing for 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.

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

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

$$\gamma(l) = E[X_t X_{t+l}] = \frac{1}{2}[(l+1)^{2H} - 2l^{2H} + |l-1|^{2H}] \quad \text{with } l \geq 0 \tag{2.1.3}$$

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.

2.1.3.3 Fractional ARIMA (0,d,0)

A FARIMA (0,d,0) process is defined as $\{X_t = \Delta^{-d} z_t, t \geq 1\}$ where z_t are independent, identically distributed normal random variables with mean zero and unit variance (i.e.$z_t \sim$ i.i.d $\ \mathrm{N}(0,1)$). $\Delta z_t = z_t - z_{t-1}$ is the differencing operator[5]. Another way to interpret $\{X_i\}$ is as a moving average:

$$X_t = \sum_{k=0}^{\infty} c_k z_{t-k}$$

where $c_k = \binom{d}{k}(-1)^k = \dfrac{\Gamma(k+d)}{\Gamma(j+1)\Gamma(d)}$ $(k \geq 1)$. Analogous to the behaviour of the autocovariance of a fGn when $0.5 < H < 1$, when $0 < d < 0.5$ the autocovariance function of a FARIMA process satisfies:

$$\gamma(l) = E[X_i X_{i+l}] \sim C_d \, l^{2d-1} \quad \text{as} \ \ l \to \infty \tag{2.1.4}$$

It can be shown that the relationship between H and d is expressed via $d = H - 0.5$ (Beran (1994)). The FARIMA (0,d,0) is a special case of the more general FARIMA(p,d,q) process (or ARFIMA(p,d,q)), a versatile parametric model family, which is the basic motivation for the development of the FIGARCH model.

[5] We can also write $\Delta z_t = (1-L)z_t$ with $L(z_t) = z_{t-1}$. The operator L is commonly referred to as the lag operator or backwards operator, and is also often denoted as B. It follows that $(1-L)^d X_t = z_t$.

From here on, to keep notations manageable, we refer to the 'fractional' ARIMA (0,d,0) simply as "FARIMA" as oppose to the more general, and more often used, term of 'fractionally integrated' ARFIMA (p,d,q). The FARIMA process is also called a fractional difference process.

2.1.3.4 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 above mentioned FARIMA (0,d,0) and can be defined by:

$$\begin{aligned} &a(L)(1-L)^d X_t = b(L)z_t \\ &z_t \sim \text{i.i.d} \quad \text{N}(0,1) \end{aligned} \tag{2.1.5}$$

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 $(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} L^k$ (this expression is known as the hyper-geometric function where the fractional parameter $-\frac{1}{2} < d < \frac{1}{2}$. When $\{X_t\}$ satisfies all these conditions we have a stationary and invertible ARFIMA process (Bollerslev and Wright (2000)).

We are able to simulate the described long-memory processes thanks to functions incorporated in softwares such as the fArma package in R (Wuertz (2013)). The following figure plots a sample path for each of these processes with $H = 0.7$ and $N = 1000$ 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 pattern of these processes' autocorrelations reflects their long-range dependence structure: they decay hyperbolically at a slow rate (up to 100 lags in our example). For the purpose of illustrating different estimators of the Hurst index, we shall focus on examining only the fGn with $H = 0.7$ in latter sections, as the estimates of other processes can be obtained in a similar manner.

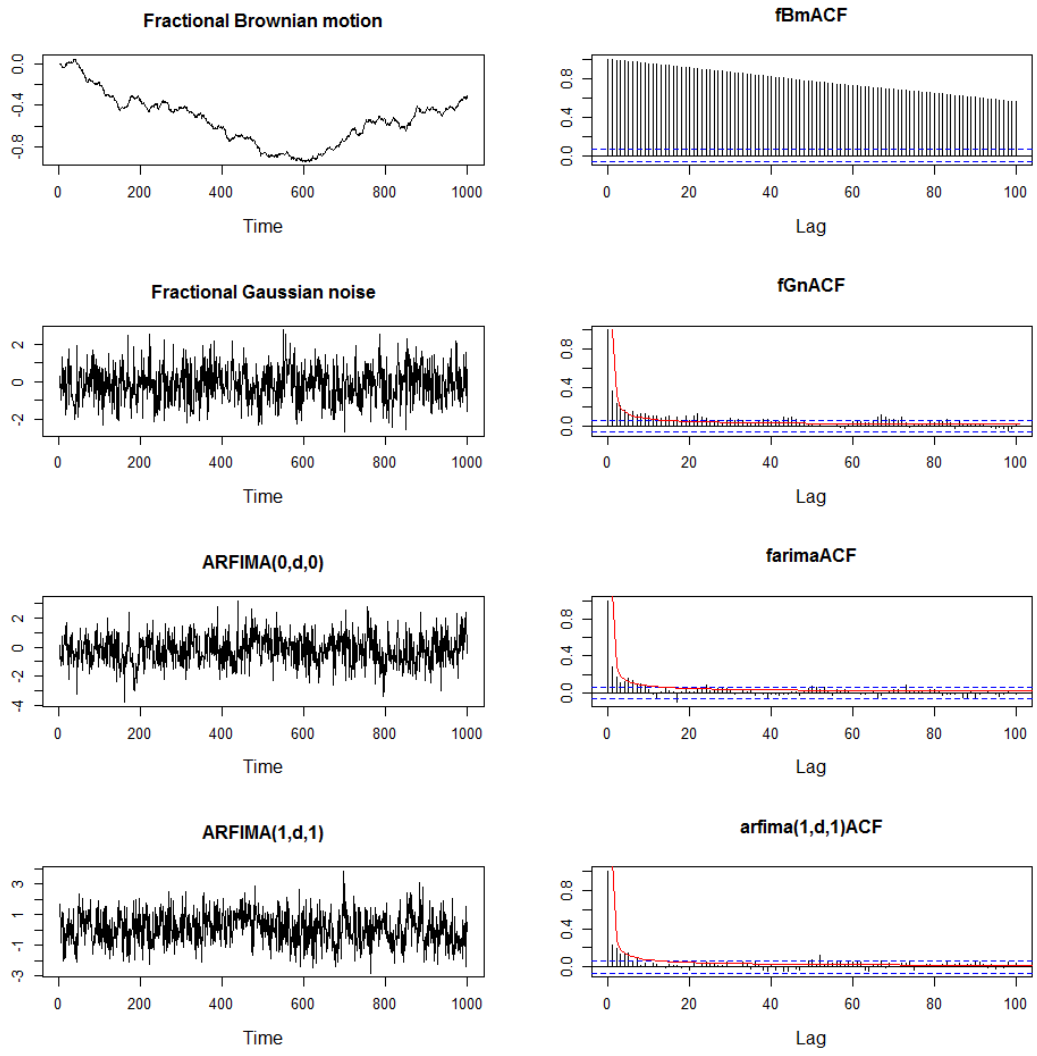


Figure 2.1 Left panel: sample paths of long memory processes with $H = 0.7$ ($d = 0.2$) and $N = 1000$. Right panel: their respective estimated and theoretical autocorrelation (or the approximate power curve in the general form of $\rho(l) \sim C|l|^{-\alpha}$). The theoretical ACF curve is indicated by a red line.
(*) Plot names indicate the abbreviations of respective processes on the left.

2.1.4 The spectral density of long memory processes

Consider the simplest case of the fractional ARIMA process class: the FARIMA(0,d,0) (also known as a Fractional difference process (FDP)) which is described in section 2.1.3 as

$$X_t = \Delta^{-d} z_t \quad \text{with} \quad z_t \sim \text{i.i.d N}(0, \sigma_z^2) \ [6]$$

[6] In some contexts the residuals z_t are standardized to be i.i.d $N(0, 1)$. Here for generality we consider z_t as a white noise process with constant variance.

This expression means that the d-th order difference of $\{X_t\}$ equals a (stationary) white noise process. A zero-mean FDP (with $-0.5 < d < 0.5$), denoted as $\{X_t\}$, is stationary and invertible (see e.g. Gençay et al. (2002) and Jensen (2000)). We have already defined its slowly decaying autocovariance function as:

$$\gamma(l) = E[X_t, X_{t+l}] \sim C_d \, |l|^{2d-1} \text{ when } l \to \infty$$

Correspondingly, for frequency f satisfying $-\frac{1}{2} < f < \frac{1}{2}$ the spectral density function of $\{X_t\}$ satisfies:

$$S(f) = \frac{\sigma_z^2}{|2 \sin(\pi f)|^{2d}} \propto f^{-2d} \text{ when } f \to 0$$

The SDFs of the FDPs with different values of d (and standard normal innovations, i.e. $\sigma_z^2 = 1$) are plotted in Figure 2.2. When $0 < d < 0.5$ (i.e. long memory exists), the slope of the SDF on a log-log scale increases as d increases. In this case, the SDF has an asymptote at frequency zero, or it is "unbounded at the origin", as Jensen (2000) put it. In other words, $S(f) \to \infty$ when $f \to 0$. Correspondingly, the ACFs decay more slowly as d increases: whereas the ACF of a process with $d \approx 0$ quickly dissipates after a small number of lags, the ACF of a process at the 'high end' of the long-memory family, with $d = 0.5$, effectively persists at all lags. The former was commonly interpreted as exhibiting 'short-memory' behaviour.

We can see the relationship between autocovariance and spectrum: the ACF decreasing towards an asymptote at very long lags, which corresponds to very low frequencies (as the observations are separated by a great time distance, i.e. the wavelength of the periodic signal becomes very high). This reminds us that the spectrum is a "representation" of the autocorrelation function in the frequency domain. In addition, Figure 2.2 shows that the higher the degree of long-memory (the higher the d), the larger the spectrum will be when $f \to 0$. It was well-established that both slowly decaying autocorrelation and unbounded spectrum at the origin independently characterize long memory behaviour (See e.g. McLeod and Hipel (1978) and Resnick (2007)). In line with these authors, Granger (1966) agrees that the pattern of power concentrating at low frequencies and exponentially declining as frequency increases, such as the ones in the top plot of Figure 2.2, is the "typical spectral shape" of an economic variable. An important remark that should be made out of this observation is that since the periodogram is very high at low frequencies, it is the low frequencies components of a long-memory process that contribute the most to the dynamics of the whole process. For our

purposes, we show in later chapters that to understand the underlying mechanism of a volatility process we might need to place more emphasis on the activities of investors with long trading horizons rather than the day-to-day, noisy activities of, say, market makers. 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 (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 popularly known "Hurst effect" in the long-memory literature. This effect is also at odd with the 'universal' wisdom of Efficient Market Hypothesis, which will be discussed below.

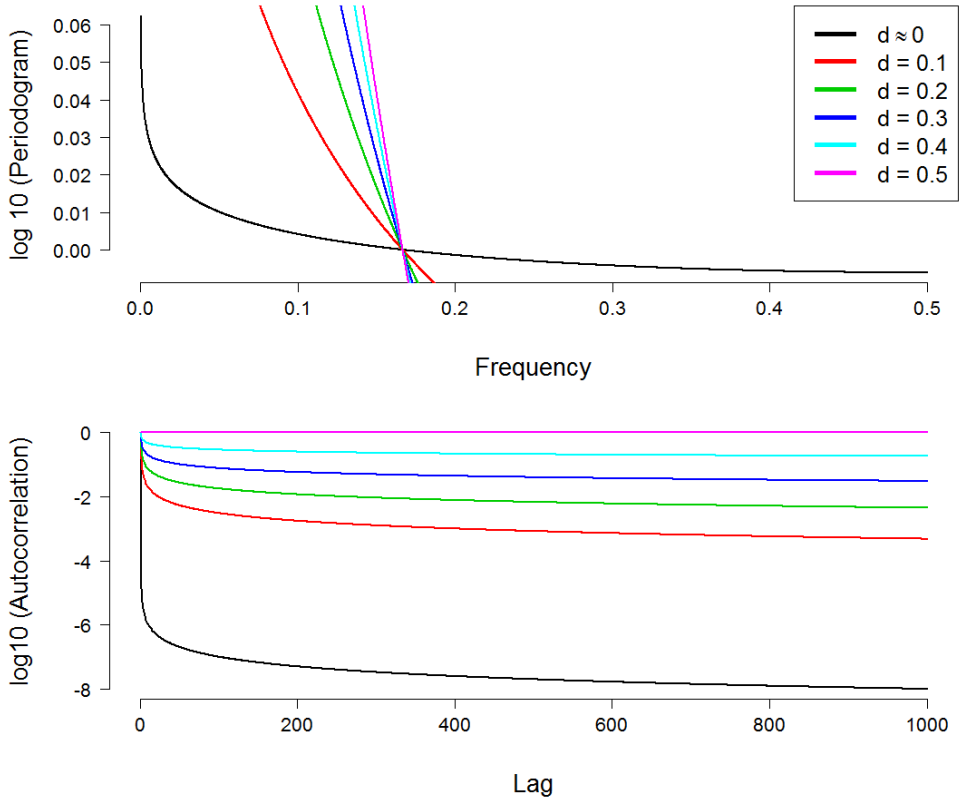


Figure 2.2 Spectral density function and ACF of a fractional difference process with various value of d. The higher the fractional difference parameter, the stronger the long-run dependence and the higher the slope of the spectrum. The rate of decay of the corresponding ACF decreases as d increases: from very quickly (when $d \approx 0$, i.e. short-memory) to effectively infinite persistence (when $d = 0.5$). Note how the asymptote at very low frequencies (in the SDF plot) is associated to the persistence at very big lags (in the ACF plot).
Note that because of scale matter, the plot of the SDFs in figure 2.2 seemingly indicates that the SDFs with $d > 0$ 'cut' the x-axis, although in fact this is not the case. All SDFs exhibit an exponential decaying pattern when plotted on appropriate scales. The higher the d, the higher the asymptote near frequency zero and the faster the SDF decays.

2.2 Methods for estimating long memory in financial time series

2.2.1 Long memory in financial time series and the contemporary arguments against market efficiency

In stock markets, analyses of long-range dependence of returns are known to yield mixed evidence [7] and 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 (which was independently developed by Samuelson (1965) and 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 attributed to a 'stylized fact' of financial volatility. It is observed that equity volatility exhibits a Hurst exponent estimated to be greater than 0.5, typically being 0.7 (Peters (1996)). 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 'ide-

[7] For instance, stock indices are documented to have displayed long-memory by Mandelbrot, B. (1971). Contrasting evidence is reported by Lo (1991) who criticized the misleading implications of the classic R/S statistic.

ally' uniform and simultaneous interpretation of information reflected in prices (which is embraced by EMH); instead it assumes that traders may decipher information in different ways and at different times [8]. If investors are influenced by events from the past, price changes are not entirely unpredictable, and the Hurst index is different from 0.5. The FMH also assumes non-normal, leptokurtic distribution of price changes, and price decreasing faster than it increases, all of which are empirically true. Intuitively, if stock prices follow a random walk/Brownian motion under the general assumption of the EMH, then its logarithmic difference (or the financial returns, as we shall discuss in the next Chapter) 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 classic EMH that caught widespread attention of academics to date is presented in the book of Lo and MacKinlay (1999).

Additionally, it should be noted that many modern financial economists generally reject the notion of a 'static' sense of market efficiency, and adopt an adaptive, evolutionary perspective instead. Indeed, it would be unreasonable to assume that efficiency can be maintained consistently, as evidenced by numerous incidents related to inefficiencies such as value stocks and small firms yielding returns higher than market average, or the various crashes over the years implying severely mispriced assets. More relevant to our study, it is the stylized clustering behaviour of stock returns and the predictability of the financial data generating process due to its inherent long-memory that may 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 underdeveloped 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, M. Hull and McGroarty (2013) use Hurst index estimates to show that different stages of emerging economies do correspond to increasing levels of efficiency and exhibit different degree 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 plays an indispensable role in maintaining efficiency itself! Along the same line, in his path-breaking article,

[8] In principle this is in line with the Mixture of Distribution Hypothesis regarding heterogeneous information arrivals discussed in section 1.1.2, Chapter 1).

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 optimizing 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 allows for "*[...] considerably 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 would be interesting then, to reconcile the trending behaviour of stock returns implied by Hurst exponent 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 capitalizing on the trending behaviour of stock price during certain periods, Mitra (2012) documented greater trading profit associated with a higher long-memory parameter for all the periods studied. On the other hand this author also observed lower profits at times when the market exhibits mean-reversion behaviour. Intuitively, technical trading strategies seeking to follow possible market 'trends' is expected to have some sort of correlation with the Hurst index, which is a reasonable measure of such trending behaviour. In our empirical study in Chapter 4 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 long-memory behaviour of financial markets, as well as some related important arguments on the trending behaviour of markets and their 'predictability', there is virtually no recognized empirical research on the topic of applying the information conveyed by the Hurst index to actually 'predict' the market movements. One possible exception is Xu and Jin (2009), who use local Hurst index estimates to forecast drastic crashes of a Chinese stock index, with relatively robust results. Ironically, in this regard, the Hurst index shares its misfortune with wavelet analysis. Despite being supported by a very strong base of theoretical background and applications in multiple natural sciences, both have gathered relatively little attention from the finance academics and consequently, from practitioners.

2.2.2 The role of Hurst index in long memory estimation

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 term dependence 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 exponent 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. This empirical law is referred to as the "Hurst effect".

The parameter H typically takes on a value in the interval $[0, 1]$ and if observations are generated from a short range dependent process then $H = 0.5$. In this case the process is said to be "self-determining". As there is no long-range dependence, time series generated by such process can not be forecast from past information. This is analogous to the case when stock prices follow a Brownian motion, with discrete realizations following a random walk model.

When $0 < H < 0.5$ we have an anti-persistent process, where past and present observations are negatively correlated. This means the behaviour of subsequent observations contradicts that of previous observations, resulting in a tendency to revert towards the mean value. Such time series exhibits the phenomenon known as "mean-reversion" in financial literature. The tendency becomes stronger as H approaches zero.

When $0.5 < H < 1$, which was the case of the annual Nile river flow time series in Hurst's original paper, we have a long-memory process. In particular, the Rescaled range analysis detected a high degree of persistence in this time series, with a Hurst index of 0.91. Variations of this type of process are too large to be explained by a pure random walk. Such processes exhibit a trending behaviour. As discussed in subsection 2.1.3, one typical time series generated by such a process is known as a fractionally integrated series, with the 'fractional' degree of integration $d = H - 0.5$. Generally, the interpretation of H and d with regards to the nature of long-memory is summarized in Table 2.1:

Hurst exponent	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 reversing
$H = 1/2$	$d = 0$	Random, Brownian motion
$1/2 < H < 1$	$0 < d < 1/2$	Long range dependence
$H \geq 1$	$d \geq 1/2$	Non stationary

Table 2.1 Categorizing stochastic processes based on their long-memory property.

2.2.3 The method of Rescaled range analysis (R/S for short)

Since the establishment of the R/S 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 generalization of the Rescaled range method were popularized by Mandelbrot and Van Ness (1968), who also coined the term Hurst index (or Hurst exponent) in recognition of Hurst. Among the first to use this method to examine long-range dependence in common stock returns is Mandelbrot. Furthermore, Mandelbrot and Wallis (1968) and Mandelbrot (1972), together with many others, radically refined R/S statistic. In particular, they advocate its robustness in detecting as well as estimating long-range dependence even for 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 presented 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-range, corresponding to high frequency autocorrelation or heteroskedasticity. There are two important implications for us 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 an AR(1)) might be adequate.

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 $\{r_t\}$ $(t = 1, 2..., n)$ divided into several 'ranges'(or 'blocks') with range size k satisfying

$1 \leq k \leq n$, the R/S statistic is:

$$Q_n = \left(\frac{R}{S}\right)_n = \frac{1}{\hat{\sigma}_n}\left[\max_{1\leq k\leq n}\sum_{t=1}^{k}(r_t - \overline{r}_n) - \min_{1\leq k\leq n}\sum_{t=1}^{k}(r_t - \overline{r}_n)\right] \tag{2.2.1}$$

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 $\overline{r}_n = \frac{1}{n}\sum_{j=1}^{n} r_t$. Because $\sum_{t=1}^{n}(r_t - \overline{r}_n) = 0$, 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 $Q_n \geq 0$.

The denominator $S_n = \hat{\sigma}_n$ is the maximum likelihood estimated standard deviation (i.e. $\hat{\sigma}_n \equiv \frac{1}{n}\sum_{t=1}^{n}[r_t - \overline{r}_n]$). 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) utilizes the square root of a consistent estimator of the variance of the cumulative sum, first proposed by Newey and West (1987). Specifically:

$$\begin{aligned} S_n^2 &= \hat{\sigma}_n^2(l) = \hat{\sigma}_n^2 + 2\sum_{j=1}^{l} w_j(l)\ \hat{\gamma}_j \\ w_j(l) &= 1 - \frac{j}{l+1} \text{ with } l < n \end{aligned} \tag{2.2.2}$$

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 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 is 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:

$$Q_u = \frac{1}{u}\sum_{u=1}^{n/k} Q_k(u) \tag{2.2.3}$$

The Hurst exponent is approximated by the slope of the regression of log Q_u against log k. The resulting 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 processes as constructed in section 2.1. The procedure described above provides a visual representation of the Hurst exponent and was also called the 'graphical' R/S method. In addition, Lo 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.

Application of this method on simulated time series with different degrees of dependence, as shown in Teverovsky, Taqqu, and Willinger (1999), indicates bias towards accepting the null hypothesis and illustrates the positive relationship between R/S statistic and the lag l given $0.5 < H < 1$. Applying this method on a simulated fGn with $H = 0.7$ gives us an estimate of 0.6126, as shown in Figure 2.3.

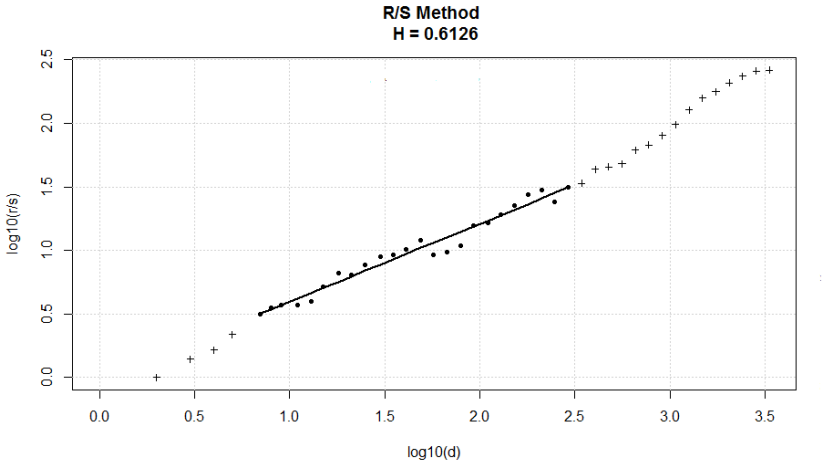


Figure 2.3 Log-log plot for estimating Hurst index from a fGn process with $H = 0.7$ by R/S method.

As implied in a counterargument raised by Willinger, Taqqu, and Teverovsky (1999), we should be cautious of the implication of Lo's modified method because of its tendency to reject long range dependence even when such behaviour exists (albeit weakly). Therefore, despite the popularity the R/S statistic has enjoyed over the years, we follow these authors' advice of relying not solely on this technique, but on a diverse range of well-established alternatives in the literature of long range dependence estimation. Appendix B provides a brief description of the

alternative methods we used to estimate the long-range dependence parameter H beside the R/S method. These methods are suggested in an empirical review of Taqqu, Teverovsky, and Willinger (1995), and were subsequently incorporated in the R package fArma. At the end of this Chapter we shall compare the performance of these methods and the wavelet-based maximum likelihood estimator to show the superiority of the latter.

2.2.4 Wavelet-based maximum likelihood estimator of long-memory

To avoid the burden of computing the exact likelihood function of their wavelet-based long memory estimator, Gençay et al. (2002) utilize an approximation to the covariance matrix obtained via a Discrete Wavelet Transform (DWT)[9]. Denoting $\{X_t\}$ as a fractional difference process with dyadic length $N = 2^J$ and covariance matrix $\sum_X$, the likelihood function is defined as:

$$L(d, \sigma_z^2|X) = (2\pi)^{N/2} \left|\sum_X\right|^{-1/2} \exp\left[-\frac{1}{2} X^T \sum_X^{-1} X\right]$$

where $\left|\sum_X\right|$ is the determinant of $\sum_X$. Also we have the approximate covariance matrix given by

$$\sum_X \approx \widehat{\sum_X} = \mathcal{W}^T \Omega_N \mathcal{W}$$

where $\mathcal{W}$ is the orthonormal matrix representing the DWT while Ω_N is a diagonal matrix include the variances of DWT coefficients computed from the process.

The approximate likelihood function and its logarithm are:

$$\begin{aligned} \widehat{L}(d, \sigma_z^2|X) &= (2\pi)^{-N/2} \left|\widehat{\sum_X}\right|^{-1/2} \exp\left[-\tfrac{1}{2} X^T \widehat{\sum_X}^{-1} X\right] \\ \log \widehat{L}(d, \sigma_z^2|X) &= -2 \log\left[\widehat{L}(d, \sigma_z^2|X)\right] - N \log(2\pi) \\ &= \log\left(\left|\widehat{\sum_X}\right|\right) + X^T \widehat{\sum_X}^{-1} X \end{aligned} \tag{2.2.4}$$

As noted earlier, Gençay et al. (2002) introduced the wavelet variance S_j for scale λ_j which satisfies $S_j(d, \sigma_z^2) = \sigma_z^2 S_j'(d)$. The properties of diagonal and orthonormal matrices allow us to rewrite the approximate log-likelihood function

[9] The definition and further discussion on DWT can be found in Appendix A.

in Equation 2.2.4 as:

$$\widehat{L}(d, \sigma_z^2|X) = N \log(\sigma_z^2) + \log[S'_{J+1}(d)] + \sum_{j=1}^{J} N_j \log[S'_j(d)] + \frac{1}{\sigma_z^2}\left[\frac{v_J^T v_J}{S'_{J+1}(d)} + \sum_{j=1}^{J} \frac{w_j^T w_j}{S'_j(d)}\right] \tag{2.2.5}$$

Noted that this maximum likelihood procedure requires us to find values of d and σ_ϵ^2 to minimize the log-likelihood function. Next we set the differentiated Equation 2.2.5 (with respect to σ_z^2) to zero and then find the MLE of σ_z^2:

$$\widehat{\sigma}_z^2 = \frac{1}{N}\left[\frac{v_J^T v_J}{S'_{J+1}(d)} + \sum_{j=1}^{J} \frac{w_j^T w_j}{S'_j(d)}\right]$$

Finally, putting this value into Equation 2.2.5 to get the reduced log-likelihood, which is a function of only the parameter d:

$$\widehat{L}(d, \sigma_z^2|X) = N \log(\widehat{\sigma}_z^2) + \log[S'_{J+1}(d)] + \sum_{j=1}^{J} N_j \log[S'_j(d)]$$

To test this method we apply it to our simulated fGn dataset with $H = 0.7$ and the volatility series of S&P500 index (proxied by daily absolute returns). Because a dyadic length signal is crucial for this experiment, we obtain daily data ranges from 06 Feb 1981 to 31 Jul 2013 (from `http://finance.yahoo.com`), for a total of $8192 = 2^{13}$ working days. Also, we set the number of simulated fGn observations equals to 8192 for comparison. We choose an LA8 wavelet with the decomposition depth set to 13 levels.

Figure 2.4 summarizes our results. Because the actual values of the SDF are very small, we substitute it with its base 10 logarithmic transform to make the plot visually clear. Estimates of d are 0.2435 and 0.2444 (corresponding to $H = 0.7435$ and $H = 0.7444$) for the simulated fGn and S&P500 daily volatility processes, respectively. Corresponding values of $\hat{\sigma}_z^2$ (or the residuals' variance) are 0.8359 and 6.2236×10^{-5}. Subsequently we have the corresponding time series models:

$$\begin{aligned} &(1-L)^{0.2435} X_t = z_t \quad \text{with} \quad z_t \sim \text{i.i.d } N(0, 0.8359) \\ &(1-L)^{0.2444} |r_t| = u_t \quad \text{with} \quad u_t \sim \text{i.i.d } N(0, 6.2236 \times 10^{-5}) \end{aligned} \tag{2.2.6}$$

To further illustrate the ability of our estimator in capturing long-memory behaviour, for each case we plot the theoretical SDF of a fractional difference process with a parameter d set to equal that of the estimated value. Then we 'fit'

this SDF (indicated by a green line) with the corresponding periodogram/spectral density function obtained from the data. In line with Gençay et al. (2002)'s findings, for both cases the two spectra are in "good agreement" in terms of overall shape save for some random variation. However, we obtain a much smaller value of $\hat{\sigma}_z^2$ for the S&P500 series, thus random variation is less severe in its case. In other words the green line approximates the spectrum of the volatility series better than in the case of the fGn. In summary, it can be concluded that this method is effective regarding detecting long range dependence. The result also indicates that the daily S&P500 volatility series can be reasonably modelled as an fGn process since the two exhibit very similar long-memory parameters.

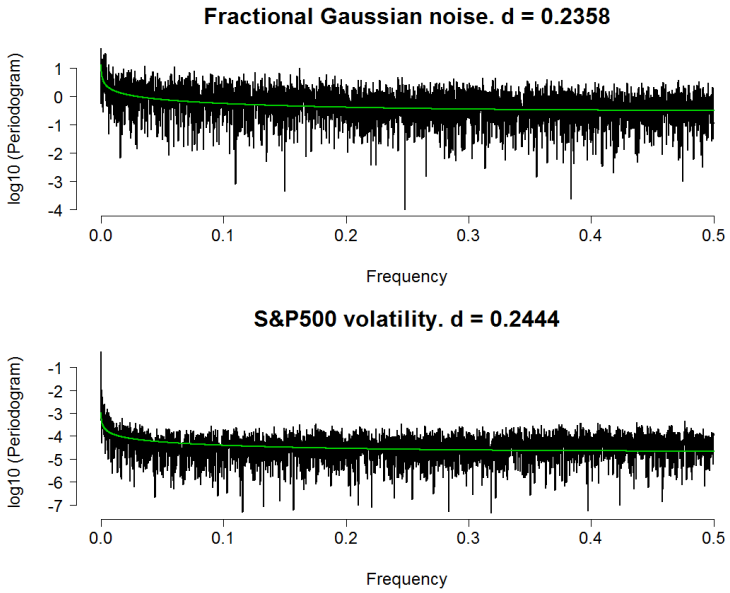


Figure 2.4 Spectral density functions (on a $\log_{10}$ scale) of a fractional Gaussian noise process and the daily volatility process of S&P500. Highest frequency being 0.5 corresponds to the Nyquist frequency. In both plots the green line indicates theoretical spectrum of long-memory process with values of d estimated by the waveMLE method.

2.3 Measurements of estimators' performance via simulation study

Now that we have introduced a plethora of estimating methods for long-memory parameter, let us compare their performance, especially against the wavelet-based MLE. We follow Taqqu et al. (1995)'s framework for comparing the performance of described methodologies: first, we simulate $N = 500$ realizations of the long-memory processes (i.e. fBm, fGn, FARIMA (0,d,0) and ARFIMA (1,d,1)); each realization has a length of 10,000 and is generated with the 'nominal' Hurst exponent set to $H = 0.7$. Then we applied all 9 estimators to each

realization, to obtain a sample of $N = 500$ estimates for each of the methods.

Next, we compute the sample mean, standard deviation and the square root of the Mean Squared Errors for each sample as follows:

$$\begin{aligned} \overline{H} &= \frac{1}{N}\left(\sum_{n=1}^{N} H_n\right) \\ \widehat{\sigma} &= \sqrt{\frac{1}{N-1}\Big[\sum_{n=1}^{N} H_n^2 - \frac{1}{N}\left(\sum_{n=1}^{N} H_n\right)^2\Big]} \\ \sqrt{MSE} &= \sqrt{\frac{1}{N}\sum_{n=1}^{N}(H_n - H)^2} \end{aligned} \tag{2.3.1}$$

where H_n is the estimate of Hurst index obtained from the n-th (simulated) realization of each process in each sample. Similar to conventional estimating techniques, here the standard error indicates the significance of the estimator while the Mean squared Error measures its performance by comparing it with the nominal value.

We repeat this procedure with $H = 0.5$ and $H = 0.9$ (approximately representing the lower and upper 'bounds' of the Hurst exponent for the stationary long-memory class). For comparative purposes, we also estimate H for a sample of 50 realizations ($N = 50$) as performed by Taqqu et al. (1995).

We noted that the MSE incorporate an indicator of bias within our simulated samples, that can be expressed as:

$$\text{MSE} = \text{Sample variance} + (\text{Bias})^2 = \widehat{\sigma}^2 + (\overline{H} - H)^2$$

so that the estimator yielding the smallest value of $\sqrt{MSE}$ is considered to perform most effectively.

The results for the simulation experiment with the fGn and FARIMA (0,d,0) (for which $\hat{H} = \hat{d} + 0.5$) [10] are reported in Table 2.2. Results for fBm and ARFIMA (p,d,q) processes are not reported, but are very similar to those reported.

[10] Because the R function used to generate simulated ARFIMA processes is not designed for $d = 0$ ($H = 0.5$) we opt to use $d = 0.0001$. Nevertheless, the generated process is essentially a short-range dependence process which suits our purpose.

		Fractional Gaussian Noise (fGn)						**FARIMA(0,d,0)**					
		H=0.5		H=0.7		H=0.9		H=0.5		H=0.7		H=0.9	
Method	Measurement	N=50	N=500	N=50	N=500	N=50	N=500	N=50	N=500	N=50	N=500	N=50	N=500
R/S	$\overline{H}$	0.564044	0.56319	0.716726	0.712847	0.822769	0.830427	0.559995	0.570928	0.682169	0.702703	0.846436	0.827914
	$\widehat{\sigma}$	0.068031	0.084093	0.094401	0.088482	0.088624	0.094284	0.08268	0.085137	0.079361	0.089118	0.109524	0.089754
	$\sqrt{MSE}$	0.092937	0.105121	0.094938	0.089366	0.116884	0.117099	0.101482	0.110746	0.080561	0.08907	0.120932	0.115048
aggVar	$\overline{H}$	0.493106	0.49756	0.685467	0.684106	0.846928	0.843435	0.494856	0.496049	0.689653	0.685929	0.844181	0.843381
	$\widehat{\sigma}$	0.024581	0.026615	0.02971	0.029432	0.033844	0.030224	0.030429	0.027867	0.028472	0.029904	0.031551	0.02869
	$\sqrt{MSE}$	0.025291	0.0267	0.032806	0.033436	0.062762	0.064119	0.03056	0.028119	0.030025	0.033022	0.063964	0.06346
diffaggVar	$\overline{H}$	0.508194	0.51835	0.714891	0.709516	0.899048	0.902468	0.515293	0.511275	0.719475	0.705657	0.900607	0.899759
	$\widehat{\sigma}$	0.064374	0.060957	0.05419	0.054835	0.058556	0.05672	0.050396	0.054595	0.056964	0.055248	0.059264	0.05583
	$\sqrt{MSE}$	0.064251	0.063601	0.055674	0.055628	0.057976	0.056717	0.05218	0.055694	0.05966	0.055482	0.058672	0.055774
AbaggVar	$\overline{H}$	0.498112	0.502188	0.69093	0.689505	0.853343	0.849268	0.497348	0.500607	0.694737	0.691525	0.849984	0.848714
	$\widehat{\sigma}$	0.026069	0.027586	0.030306	0.030654	0.033511	0.032466	0.031187	0.029149	0.030723	0.031045	0.033459	0.031593
	$\sqrt{MSE}$	0.025876	0.027645	0.031342	0.032387	0.057249	0.060214	0.030987	0.029126	0.030867	0.032151	0.059989	0.06022
Per	$\overline{H}$	0.497351	0.501328	0.704646	0.706274	0.911928	0.911946	0.498027	0.499385	0.705351	0.703871	0.908388	0.908864
	$\widehat{\sigma}$	0.018536	0.021702	0.019872	0.020392	0.020889	0.020556	0.019497	0.021525	0.019393	0.021554	0.021267	0.021134
	$\sqrt{MSE}$	0.01854	0.021721	0.020214	0.021326	0.023873	0.023757	0.019402	0.021512	0.01993	0.021878	0.022663	0.022898
modPer	$\overline{H}$	0.450786	0.456267	0.661877	0.663776	0.869063	0.869597	0.452181	0.453851	0.660308	0.658264	0.862619	0.860428
	$\widehat{\sigma}$	0.022872	0.022434	0.02436	0.021567	0.024411	0.021627	0.020106	0.023191	0.023445	0.022493	0.022643	0.021936
	$\sqrt{MSE}$	0.054172	0.049141	0.04511	0.042152	0.039256	0.037298	0.051796	0.051638	0.04598	0.047401	0.043587	0.045235
Peng	$\overline{H}$	0.489161	0.49098	0.687735	0.687535	0.886076	0.885492	0.486886	0.488775	0.676324	0.676613	0.872576	0.873507
	$\widehat{\sigma}$	0.014573	0.012173	0.015251	0.014977	0.01755	0.016267	0.013205	0.011949	0.015123	0.015777	0.016769	0.017056
	$\sqrt{MSE}$	0.018045	0.015141	0.019452	0.01948	0.022265	0.021785	0.018517	0.016386	0.028012	0.028202	0.032057	0.031499
Higuchi	$\overline{H}$	0.474195	0.478265	0.67017	0.669771	0.860494	0.857328	0.473408	0.475953	0.674804	0.670161	0.864872	0.85892
	$\widehat{\sigma}$	0.017094	0.019349	0.025276	0.024767	0.036007	0.041954	0.022524	0.020202	0.029415	0.026236	0.043604	0.041251
	$\sqrt{MSE}$	0.030859	0.029087	0.038935	0.039072	0.05321	0.059812	0.034703	0.031394	0.038506	0.039716	0.055653	0.058188
waveMLE	$\overline{H}$	0.497214	0.500651	0.725488	0.726317	0.915463	0.914725	0.500159	0.500303	0.69438	0.694159	0.89518	0.891318
	$\widehat{\sigma}$	0.007833	0.008072	0.007825	0.007873	0.016546	0.016894	0.009373	0.008198	0.009471	0.00843	0.023551	0.021036
	$\sqrt{MSE}$	0.008239	0.00809	0.026639	0.027468	0.022525	0.022398	0.00928	0.008196	0.010931	0.010249	0.023807	0.022738

Table 2.2 Performance comparison of different long-memory estimators. Nomenclatures are as follows: (1) R/S : Rescaled range; (2) aggVar: Aggregated variance; (3) diffaggVar : differenced variance; (4) AbaggVar: absolute moments; (5) Per: Periodogram; (6) modPer: Modified periodogram; (7) Peng: regression of residuals; (8) Higuchi: Higuchi's method; (9) waveMLE: wavelet-based Maximum likelihood.
The methods other than R/S and waveMLE are discussed in Appendix B.

There are several observations that can be made from the above table:

- All of the proposed methods seem to estimate parameter H reasonably well, in that they can detect the correct degree of dependence structure exhibited by the simulated time series, with relatively small standard errors.
- The values of $\overline{H}$, $\hat{\sigma}$ and $\sqrt{MSE}$ do not vary significantly between the cases of $N = 50$ and $N = 500$.
- The Rescaled range method yields the least desirable performance in all simulated experiments. This is in contrast to many previous studies supporting the use of the method, yet it is in line with the view of sceptics such as Taqqu et al. (1995).

Assessing the potential of waveMLE When looking more closely at Table 2.2 we can see that the Wavelet-based MLE performs relatively robustly compared to several other methods. In particular, in the case of the simulated fGn, when $H = 0.5$ (when the process becomes a Bm) waveMLE is in absolute superiority. For a 'typical' long-memory process ($H = 0.7$) waveMLE ranks third, and for process exhibiting extreme long-range dependence behaviour ($H = 0.9$) this estimator ranks second. When it comes to the FARIMA(0,d,0) process, waveMLE performs best in all cases. We illustrate these arguments by establishing a ranking system based on $\sqrt{MSE}$ (with $N = 500$) and present it in Table 2.3.

Additionally, when $H = 0.5$ and $H = 0.7$ (corresponding to the value generally expected to be observed in financial returns and financial volatility time series, respectively) waveMLE provides the smallest value of $\sqrt{MSE}$ on average, which is also smaller than those obtained when estimating $H = 0.9$ (which is unlikely to be observed). This feature is crucial to our empirical study presented in Chapter 4.

Furthermore, in cases where waveMLE is not the best estimator, the difference between its performance and that of the best estimator is not significant. For example, in the case of fGn with $H = 0.7$ and $H = 0.9$ this difference is only measured by units of 0.1%. To see how important this is, consider the performance of Peng method (which outperforms waveMLE in these cases): when Peng is not the best, the difference between its performance and the best estimator (waveMLE) is in units of 1%. Colloquially speaking, waveMLE is rarely beaten by another estimator, and when it is, it does not get beaten by a large margin. Nevertheless, with all evidence in clear favour of waveMLE, in practice we still need keeping in mind the main limitation of this seemingly superior estimator,

that is, it can only be applied on a dyadic length time series.

Estimators ranking based on Square root of MSE					
Fractional Gaussian Noise (fGn)					
Method	**H=0.5**	**Method**	**H=0.7**	**Method**	**H=0.9**
waveMLE	0.00809	Peng	0.01948	Peng	0.0217846
Peng	0.015141	Per	0.021326	waveMLE	0.0223977
Per	0.021721	waveMLE	0.027468	Per	0.0237573
aggVar	0.0267	AbaggVar	0.032387	modPer	0.0372976
AbaggVar	0.027645	aggVar	0.033436	diffaggVar	0.0567174
Higuchi	0.029087	Higuchi	0.039072	Higuchi	0.0598122
modPer	0.049141	modPer	0.042152	AbaggVar	0.0602138
diffaggVar	0.063601	diffaggVar	0.055628	aggVar	0.0641194
R/S	0.105121	R/S	0.089366	R/S	0.1170989
FARIMA(0,d,0)					
Method	**H=0.5**	**Method**	**H=0.7**	**Method**	**H=0.9**
waveMLE	0.008196	waveMLE	0.010249	waveMLE	0.0227378
Peng	0.016386	Per	0.021878	Per	0.0228976
Per	0.021512	Peng	0.028202	Peng	0.0314989
aggVar	0.028119	AbaggVar	0.032151	modPer	0.0452348
AbaggVar	0.029126	aggVar	0.033022	diffaggVar	0.0557743
Higuchi	0.031394	Higuchi	0.039716	Higuchi	0.0581878
modPer	0.051638	modPer	0.047401	AbaggVar	0.0602195
diffaggVar	0.055694	diffaggVar	0.055482	aggVar	0.0634601
R/S	0.110746	R/S	0.08907	R/S	0.1150482

Table 2.3 Performance ranking of different estimators. The methods are ordered based on increasing values of $\sqrt{MSE}$. This statistic is calculated from a sample of 500 estimates for each method. Nomenclatures are similar to those in Table 2.2.

Chapter 3

Volatility modelling with GARCH

Preamble

In this Chapter we examine the capability of a model family known as Generalized Autoregressive Conditional Heteroskedastic (GARCH) in providing reliable volatility estimates. The Chapter is organized as follows: Section 3.1 and 3.2 provide an introduction to volatility and discuss how several 'stylized facts' of volatility were intuitively incorporated into various specifications of GARCH. Sections 3.3 illustrates the goodness-of-fit of the GARCH models in-sample via various diagnostic tests. Section 3.4 demonstrates the improved performance of GARCH out-of-sample when we use a volatility measure constructed from high-frequency data. In addition, new evidence from multi-scale analysis suggests our wavelet decomposition may provide a better way of approximating the true volatility factor.

3.1 A primer on volatility

Poon and Granger (2003) emphasis the role of volatility in modern finance literature. The importance of volatility is apparent in a large variety of disciplines, of which the following are but a few: derivative pricing (e.g. option pricing depends heavily on underlying stock's volatility); risk management (volatility forecasting has become crucial to determining Value-at-risk especially after the global implementation of the first Basel Accord in 1988); policy making (which relies on estimates of financial vulnerability via proxies such as volatility).

Understanding the nature of volatility is therefore becoming ever more crucial in financial modelling. Taylor (2005) proposed one of the most popular definitions of this variable: it is defined by the variance (or standard deviation in some contexts) of a historical set of returns $\{r_t | t = 1, \ldots, T\}$ whose mean is $\bar{r} = \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t$:

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(r_t - \bar{r})^2$$

Simply put, volatility of an asset is the standard deviation of the returns on that asset over a period of time (J. Hull (2006)). In the finance literature, volatility is commonly measured as the absolute value of the returns. Though statistically improper, absolute value provides the best estimate of the standard deviation of a single return at a single period (Taylor (2005)). As volatility captures the variability in asset prices, it is generally taken as a proxy for risk. Poon and Granger (2003) examine volatility via the continuous time martingale that generates instantaneous returns:

$$d(\ln p_t) = \sigma_t \, dW_{p,t}$$

where p_t is the price and $dW_{p,t}$ denotes a standard Wiener process. In this context, volatility can be thought of as a "scale" parameter which adjusts the size of the variation associated with a stochastic Wiener process. Then, the conditional variance of this return over the period $[t, t+1]$ is called the "integrated volatility": $\int_0^1 \sigma^2_{t+\tau} d_\tau$. Fundamental derivatives pricing theories were built upon this quantity under stochastic volatility (See for example Melino (1991), J. Hull and White (1996) and J. Hull (2006)). In general, σ_t is unobservable. However, if we can obtain a number of continuously compounded returns over one time unit, the sum of the squared value of these returns shall provide us with a consistent estimate of the integrated volatility. This estimate is known as "realized volatility". Poon and Granger (2003) show that this estimate becomes more accurate when the number of returns computed at higher frequency (or the sampling rate) increases. Proof of this proposition can also be found in Barndorff-Nielsen and Shephard (2002). This means that given high-frequency data availability, the latent volatility is theoretically observable. Beside being an "error-free" volatility estimator, realized volatility is relatively simple to compute. A reasonable number of studies illustrate the importance of volatility measures constructed from returns recorded at high frequencies (See Andersen and Bollerslev (1998) and Andersen et al. (1999), among others).

3.2 Modelling volatility with the GARCH family

3.2.1 The original ARCH and GARCH(1,1)

The foundation of a model that could effectively capture all the stylized facts of returns has attracted enormous attention over time. In the time domain, there

are two classes of volatility models that enjoy vast popularity among academics, namely: the GARCH family models and the stochastic volatility models. While the former studies volatility as a function of observables, the latter incorporates not only observables, but also unobservable volatility components (See Taylor (2005) for a detailed review of the topic). For our purposes, it is sufficient to study the first branch, which originates from the introduction of the Autoregressive Conditional Heteroskedastic (ARCH) model first proposed in the seminal paper of Engle (1982). This model is designed to capture the time-varying nature of conditional volatility given the historic information of returns. It is important to note that stylized facts of volatility serve as theoretical motivations of specific extensions of the ARCH model. The paper by Engle and Patton (2001) reviews major stylized facts of volatility that should be incorporated in a good model. Specifically, these facts include: (i) volatility clusters, (ii) persistence, (iii) mean reversion and (iv) asymmetric impacts of innovations. Unable to cover the enormous body of research on this model family, our paper only investigates the frameworks most directly designed to incorporate these four stylized facts. Next we shall discuss the salient specifications and their relevance to our research in more detail.

Following the traditional approach we define the one-period continuously compounded returns as the first order difference of the natural logarithm of a discrete price sequence. As Taylor (2005) argued, the number of dividend payment days is very small compared to the sample of trading days and thus dividend can be ignored in this formula:

$$r_t = \ln(p_t) - \ln(p_{t-1}) = \Delta\ln(p_t)$$

Assuming asset returns are generated by the following process [11]:

$$r_t = \mu_t + \sigma_t z_t \quad \text{where} \quad z_t \sim \text{i.i.d} \quad \text{N(0,1)}$$

in which $\mu_{t|t-1} = E_{t-1}[r_t]$ and $\sigma_{t|t-1} = E_{t-1}[(r_t - \mu_t)^2]$ are the process' conditional mean and variance respectively, given the information set at time $t-1$ (which is the set of returns up to $t-1$, inclusive).

We also specify the 'residual returns' as $u_t = r_t - \mu_t = \sigma_t z_t$. Note that the conditional mean μ_t tends to be modelled as an ARMA process in some contexts in order to free the residuals of serial correlations. Nevertheless this variable is

[11] This is the standard specification of all ARCH-type models examined in this paper. Therefore we implicitly assume it in most of the later discussions.

usually set to zero or constant [12], as over the long term, in a large sample the average returns tend to be insignificant (Taylor (2005)) [13] . The (population) unconditional mean and variance can be defined as:

$$\mu = E[r_t] \text{ and } \sigma^2 = E[(r_t - \mu)^2]$$

As its name suggests, ARCH models conditional volatility auto-regressively, meaning future volatility is specified to be linearly dependent on immediate previous squared residual returns. In other words, large/small deviations from mean return is followed by high/low subsequent variances. This allows us to effectively capture volatility clustering feature and is one of the reasons for the original success of the ARCH-type model:

$$\sigma_t^2 = \omega + \alpha(u_{t-1}^2) \tag{3.2.1}$$

with $\omega > 0, \alpha > 0$. This is the simplest form of ARCH: the ARCH(1) model. When we increase the number of autoregressive terms to p, we have its extended form: the ARCH(p) model [14]:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i u_{t-i}^2$$

As Campbell, Lo, and MacKinlay (1997) pointed out, setting volatility to be time-varying allows us to capture basic stylized facts of financial returns, most notably the property that returns' distribution is heavy tailed with excess kurtosis.

In this case the returns process is stationary only if $\sum_{i=1}^{p} \alpha_i < 1$ $(\alpha_i > 0)$ so that it has a finite long term unconditional variance, which can be computed as $\sigma^2 = \text{Var}(u_t) = \dfrac{\omega}{1 - \sum_{i=1}^{p} \alpha_i}$. Also, the ARCH process is then mean-reverting.

[12] In general there are three specifications for conditional mean μ_t: we can treat it as zero, a constant, or adding an ARCH term determinant (meaning the expected value of returns also depends on past evaluation of volatility). The last type is commonly referred to as the ARCH-in-mean or ARCH-M model. The difference only results in different forecast of returns while not affecting our main interest - the volatility estimates - in any significant way.

[13] Which is why the residual returns (or 'de-meaned' returns) and original returns series have similar properties and the two are interchangeably used in different GARCH specifications. From our perspective these processes also have the same conditional variance.

[14] p is also known as the lag parameter in this context.

GARCH (1,1) : Despite its inherent ability to capture volatility clustering, ARCH model fails to account for volatility persistence, as the clusters are short-lived without accounting for extra lagged u_t terms. Bollerslev (1986) augmented this model by adding a lagged term of conditional volatility, forming the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) (1,1) model [15]:

$$\sigma_t^2 = \omega + \alpha(u_{t-1}^2) + \beta\sigma_{t-1}^2 \tag{3.2.2}$$

where $\omega > 0, \alpha > 0, \beta > 0$ and $\alpha + \beta < 1$. It can be shown that the GARCH(1,1) is analogous to an ARCH(∞) [16]. The GARCH(1,1) inherits its predecessor's ability to capture the excess kurtosis of returns' distribution. Furthermore, the most important application of this model is its ability to forecast future volatility. We have the l-ahead volatility forecast specified via:

$$\hat{\sigma}_{t+l}^2 = \sigma^2 + (\alpha + \beta)^l(\sigma_t^2 - \sigma^2)$$

Here the term $(\alpha + \beta)$ is known as "persistence parameter" as it determines the rate at which the variance forecast converges to the long-run unconditional variance $\sigma^2 = E[u_t^2] = \dfrac{\omega}{1 - (\alpha + \beta)}$ when the forecast horizon tends to infinity ($l \to \infty$). To describe this particular implication of GARCH, Engle (2001) stated that "*the GARCH models are mean reverting and conditionally heteroskedastic, but have a constant unconditional variance*"(p.160). This is an important complement of GARCH to ARCH. A related quantity specifying the degree of volatility persistence is the "half-life", denoted as k. It is the time needed for the variance to move halfway towards its unconditional level. Generally we have $(\alpha+\beta)^k = 0.5$ or $k = \ln(0.5)/\ln(\alpha + \beta)$. For the GARCH model to be stationary we must have $(\alpha + \beta) < 1$. We can write the generalized form of this model as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{3.2.3}$$

[15] Also, ARCH model is analogous to an AR model on squared residuals, whilst GARCH can be thought of as an ARMA model on squared residuals.

[16] By recursively replacing the past volatility terms with their GARCH forms we can write the GARCH model as:

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha u_{t-1}^2 + \beta\sigma_{t-1}^2 \\ &= \omega + \alpha u_{t-1}^2 + \beta(\omega + \alpha u_{t-2}^2 + \beta\sigma_{t-2}^2) = \cdots = \frac{\omega}{1-\beta} + \alpha/\beta \sum_{j=1}^{\infty} \beta^j u_{t-j}^2\end{aligned}$$

So that it becomes an ARCH(∞) model.

For the sake of tractability we only focus on the simple GARCH(1,1) as well as its related models with the same ARMA order from here on. Many researchers strongly suggest the comparability of such simple models to the ones incorporating higher order ARMA terms. Furthermore, it is widely documented that GARCH(1,1) generally equals or rivals other, more complex models in the same family, especially in terms of out-of-sample forecast performance (See Hansen and Lunde (2005)). Additionally, we present reviews of the most popular extensions of the standard GARCH model in Appendix C. These alternatives include IGARCH, FIGARCH, EGARCH and GJR-GARCH and shall be used in combination with the standard GARCH, as this is a conventional approach for estimating volatility.

Estimating GARCH with Maximum Likelihood method We know that time-varying conditional variance of residual returns, or heteroskedasticity, is the key assumption of ARCH/GARCH model. Therefore estimating this model will require other techniques than the traditional least squares approach, since the latter assume constant expected value of squared errors, or homoskedasticity [17]. A well-established alternative approach used in GARCH is the Maximum likelihood estimate, in which a likelihood function given the dataset of returns ($L = L(\Theta; r_1, r_2, \ldots, r_T | I_0)$) is maximized to find the estimated parameters ($\Theta = \{\omega, \alpha, \beta, \mu\}$). I_0 is the initial condition used to start the recursion for conditional mean and variance functions. Here we opt for using the sample mean of squared residuals as initial condition, as suggested by Bollerslev and Mikkelsen (1996). Assuming returns are normally distributed and uncorrelated, we can write the likelihood function of GARCH(1,1) as product of marginal density functions (Reider (2009)):

$$\begin{aligned} L &= \prod_{t=1}^{T} f(r_t | r_1, r_2, \ldots, r_{t-1}) \\ &= \frac{1}{\sqrt{2\pi\sigma_T^2}} \exp\left[\frac{-(u_T^2)}{2\sigma_T^2}\right] \frac{1}{\sqrt{2\pi\sigma_{T-1}^2}} \exp\left[\frac{-(u_{T-1}^2)}{2\sigma_{T-1}^2}\right] \ldots \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[\frac{-(u_1^2)}{2\sigma_1^2}\right] \end{aligned} \tag{3.2.4}$$

To find the estimates of the vector Θ we need to maximize the log-likehood function:

$$\ln L = -\frac{T}{2} \ln (2\pi) - \frac{1}{2} \sum_{i=1}^{T} \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^{T} \left(\frac{u_i^2}{\sigma_i^2}\right) \tag{3.2.5}$$

A similar technique is applied to other variations of GARCH mentioned in this chapter. The likelihood function will be adjusted to accommodate new vec-

[17] Engle (2001) noted that when this assumption is violated, least squares regression still yields unbiased coefficients but with artificially small standard errors.

tors of parameters corresponding to each model. Exact formulation of the Maximum likelihood estimators (MLE) and Quasi-Maximum likelihood estimators (QMLE-designed to account for non-normal distribution of the standardized innovations (returns)) for the models described below can be found in Bollerslev and Wooldridge (1992) and Taylor (2005), among others.

Non-normal conditional returns The assumption of normal distribution of returns conditional upon historical information is essential for establishing a starting point to studying GARCH models. However as empirical returns tend to exhibit excess kurtosis we need to relax this assumption for better modelling. The most widely used alternative returns' distribution are the Students t and the Generalized error distribution (GED). We shall impose in turn all three distributional assumptions for our GARCH models and compare their estimate performances.

Robust estimates Bollerslev and Wooldridge (1992) pointed out that the QMLE method provides consistent estimators for models which jointly parameterize conditional mean and variance such as GARCH. They also derive computable formulas for asymptotic standard errors that are robust to non-normality, even though normality is assumed when maximizing the log-likelihood function. Based on this important finding, we are able to provide robust estimates in our report.

3.2.2 Model diagnostic tests

Residual test A good specification would result in no serial correlation among its standardized residuals $\frac{u_t}{\widehat{\sigma}_t} = z_t$ (for the mean model) and squared standardized residuals z_t^2 (for the variance model). To test for serial correlation we employ the Portmanteau Q-statistic suggested by Ljung and Box (1978):

$$Q(L) = N(N+2)\sum_{l=1}^{L}\frac{\rho_l^2}{N-l}$$

where N is sample size, ρ_l is the sample autocorrelation function of model residuals at lag l. This test examines the null hypothesis of $H_0 : \rho_1 = \rho_2 = \cdots = \rho_L$ or the autocorrelations up to lag L are jointly insignificant.

Sign and size bias test We know that a negative and a positive return shock with the same magnitude most likely will have different impact on future volatility. As such, a GARCH model which treats the shocks symmetrically

(regardless of their signs) will tend to provide downward/upward biased volatility predictions following negative/positive shocks. This observation motivates the diagnostic tests specified by Engle and Ng (1993) which examine and distinguish the magnitude and direction of the changes in past returns. In particular, we have three separate tests for impacts of three 'factors': sign of returns; size of negative returns and size of positive returns. Furthermore, we can test for the joint effect of these factors using a regression equation presented in Engle and Ng (1993). In general, if the coefficient corresponding to a factor is statistically significant, it is an indicator that the model failed to capture the impact of that factor.

3.3 Fitting GARCH models using in-sample data

We continue this Chapter with analyses of GARCH frameworks covered in the previous section as well as in Appendix C, using the daily returns of Citigroup Inc. computed from the time series of closing prices, which are adjusted for stock splits and dividends. We use 7341 observations from 01 Dec 1977 to 31 Dec 2006 and refer to these as our in-sample data. More detailed description accompanying a larger data set of this company will be provided in Chapter 4. For the moment we only use this data set to investigate the performance of GARCH models, rather than to conduct a complete empirical study.

To illustrate the goodness-of-fit of GARCH(1,1), we plot the fitted conditional volatility σ_t (formulated in previous sections) against in-sample absolute returns in Figure 3.1. As can be seen, the overall shape of $\widehat{\sigma}_t^2$ tracks that of the actual data quite reasonably, save for the extreme values observed at several market crashes.

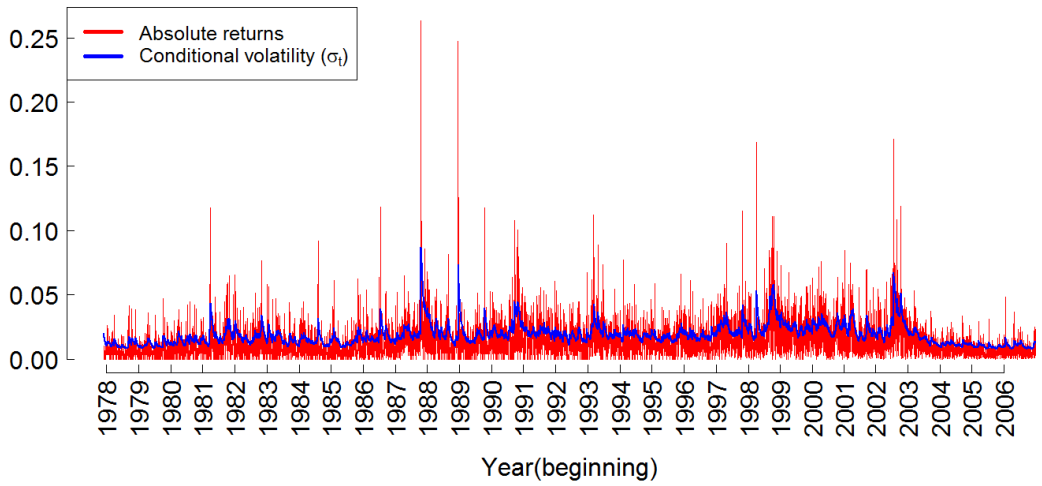


Figure 3.1 Fitting GARCH(1,1) model. In-sample data range from 01 Dec 1977 to 31 Dec 2006.

3.3.1 Comparing models' performance

Model estimates and Diagnostic statistics are reported in Table 3.1 [18]. Most of the estimated parameters are significant. The portmanteau Ljung-Box tests (up to lag 21) suggests no evidence of serial correlation among model residuals, indicating that these models seem to perform adequately. Interestingly, the EGARCH(1,1) yields the highest value of Log-likehood function compared to other GARCH models previously discussed, as well as the lowest value of Akaike Information Criterion (AIC) [19]. Both of these indicators imply better performance of EGARCH. It should be noted that the FIGARCH(1,1) ranks second in terms of performance and yield a estimated $d = 0.4221$ (corresponding to a Hurst index of 0.9221) which is significant at the 1% level. Taken together, the improved performance of EGARCH and FIGARCH imply that our volatility series does exhibit (i) a high degree of long-memory and (ii) an asymmetric relationship with positive and negative returns.

Inspired by these findings, we opt to specify yet another class of GARCH model designed to incorporate these two properties of financial volatility. The new model is called the Fractionally Integrated Exponential ARCH (FIEGARCH), which was first proposed in Nelson (1991) and then further examined by Bollerslev

[18] All estimates are obtained with the help of R package rugarch (Ghalanos (2013)), except for the results of FIGARCH and FIEGARCH (which are obtained using OXMetrics (Laurent and Peters (2002)).

[19] In its general form the AIC is defined as $:AIC = 2k - 2\ln(L)$ where k is the number of model parameters and L is the maximized likelihood value of the estimated model (Akaike (1974)). For comparison, the model with the lowest AIC is chosen.

and Mikkelsen (1996). Similar to the nesting of GARCH under FIGARCH, by incorporating the fractional orders of integration to EGARCH, we effectively nest it under the FIEGARCH model which, by definition, also nests FIGARCH. In particular, from the specification of EGARCH in Equation (C.3.1) we factorize the AR polynomial as $[1 - \varphi(L)] = \phi(L)(1 - L)^d$ and have:

$$\begin{aligned} &\ln\left(\sigma_t^2\right) = \omega + \phi(L)^{-1}(1-L)^{-d}[1 + \psi(L)]g(z_{t-1}) \\ &\text{with} \quad g(z_{t-1}) = \vartheta z_{t-1} + \gamma\left(|z_{t-1}| - E[|z_{t-1}|]\right) \end{aligned} \tag{3.3.1}$$

Analogous to the EGARCH model, here $\vartheta < 0$ and $\gamma > 0$ generally confirm the asymmetric impact of returns on future volatility. Apparently the superiority of FIEGARCH compared to EGARCH is the ability to account for long-memory, which is represented via the parameter d. When $d = 0$ this model reverts to conventional EGARCH. Bollerslev and Mikkelsen (1996) point out that the FIEGARCH process does not have to satisfy any non-negativity constraints to be well-defined. Fitting this model gives us a Log-likelihood 19281.7, which is even higher than that obtained with EGARCH. The fractional difference parameter is estimated at 0.58, which is significantly greater than 0.5 and is questionable.

Finally, to quantify the model goodness-of-fit we simply regress the absolute returns against the fitted $\hat{\sigma}_t$ via the following Equation:

$$|r_t| = a + b \;.\; \hat{\sigma}_t + u_t \tag{3.3.2}$$

and obtain the corresponding coefficients of determination: $R^2 = 0.12042$, 0.12047, 0.13096, 0.1215, 0.1301 and 0.1316 for GARCH, IGARCH, EGARCH, GJR-GARCH, FIGARCH and FIEGARCH, respectively. The estimated conditional volatility series acquired from other models do not seem to be significantly different from that of GARCH(1,1). Additionally, the plot of fitted conditional volatility of GARCH(1,1) is not very different from that of other models, given the similarity in terms of parameter estimates. The surprisingly low in-sample R^2 is a result of persistent underestimation of the fitted $\hat{\sigma}_t^2$ compared to daily volatility proxied by absolute returns (see Figure 3.1). This, in turn, may be a result of the presence of extreme volatility accompanying crises related to the 1980s housing bubble, the 1987 market crash, the 2000s Internet bubble burst or the 2008 global financial distress, for which GARCH failed to provide good estimates. In Chapter 4 we show that excluding these 'outliers' helps improve the overall goodness-of-fit.

Model estimates with Gaussian distribution for residuals												
Data: Citigroup daily returns Range: 01/12/1977 - 31/12/2006 Number of observations: 7341												
Parameters	GARCH (1,1)		IGARCH (1,1)		EGARCH (1,1)		GJR-GARCH (1,1)		FIGARCH (1,1)		FIEGARCH (1,1)	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
μ	0.000585**		0.000585***		0.000591**		0.000535**		0.000559**		0.000685****	
	(0.000178)	0.001015	(0.000175)	0.000857	(0.000184)	0.00134	(0.000172)	0.001908	(0.000179)	0.0019	(0.000166)	0.0000
ω	0.000003*		0.000003*		-0.120987***		0.000004*		0.000009****		-0.000723****	
	(0.000002)	0.02667	(0.000002)	0.036831	(0.035901)	0.000752	(0.000002)	0.044130	(0.000005)	0.0000	(0.000076)	0.0000
α	0.0853****		0.085866				0.077213****		0.3768***			
	(0.02033)	0.0000	(0.019912)	0.000016			(0.017713)	0.000013	(0.10057)	0.0002		
d									0.4221****		0.5769****	
									(0.0631)	0.0000	(0.0436)	0.0000
α^*							0.028980				0.10845	
							(0.026205)	0.26874			(0.21856)	0.6198
ϑ					-0.025562						-0.04273	
					(0.013324)	0.05561					(0.02084)	0.0404
β	0.913695****		0.914134		0.983537****		0.907297****		0.652333		0.20311	
	(0.019052)	0.0000	N/A	N/A	(0.004791)	0.0000	(0.023546)	0.0000	(0.11091)	0.0000	(0.22586)	0.3685
γ					0.176913****						0.25170	
					(0.030585)	0.0000					(0.04717)	0.0000
Diagnostic information												
AIC	-5.2246		-5.2250		-5.2431		-5.2254		-5.2360		-5.2512	
BIC	-5.2209		-5.2222		-5.2384		-5.2207		-5.2313		-5.2446	
Log Likelihood	19181.01		19181.35		19249.95		19184.87		19233.81		19281.7	
Q-stat (lag p+q+5)	21.89***	0.000548	21.85***	0.000558	21.37***	0.0006909	22.30***	0.0004582	23.16***	0.000314	22.69***	0.000386
Sign bias	0.6099	0.5419	0.6104	0.5416	0.2613	0.5344	0.5294	0.5965	0.4764	0.6337	0.4328	0.6651
Neg. size bias	1.3606	0.1737	1.3329	0.1826	1.4047	0.1602	0.9419	0.3463	0.5576	0.5770	0.23172	0.8167
Pos. size bias	0.4334	0.6647	0.4564	0.6481	0.0199	0.9841	0.3285	0.7426	0.9455	0.3444	0.3697	0.7116
Joint Effect	2.1293	0.5460	2.0754	0.5569	1.9750	0.5776	1.0012	0.8010	1.3420	0.6921	0.2233	0.9737

Table 3.1 Summary of different GARCH family model estimates. Non-normality robust standard errors are reported in parentheses.
N/A indicates where no convergence was found for the corresponding statistic.
Statistical significance indicators: * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$). All models assume Gaussian distribution of innovations.

3.3.2 Model checking via residual tests

Specifically, for GARCH(1,1) the Ljung-Box test for serial correlation yields $Q(21) = 30.693$ (p-value = 0.059) for residuals (z_t) and $Q(21) = 6.8541$ (p-value = 0.997) for squared residuals (z_t^2). This means we cannot reject the null hypothesis of no serial correlation for either of these series at 5% level for even as far as 21 lags. Indeed, the pattern exhibited by the time series plot of z_t in Figure 3.2 is analogous to that of a white noise process, and an indicator of the appropriateness of our GARCH model (for similar analyses, see Tsay (2001), p.97):

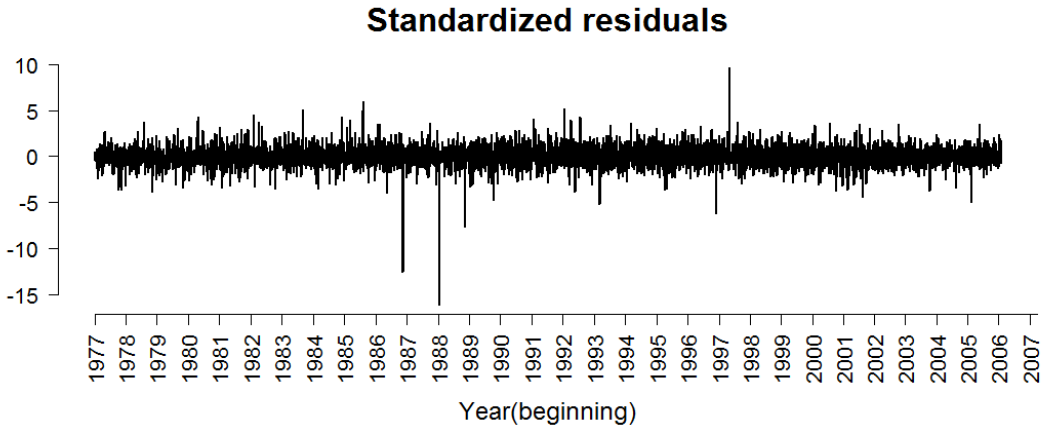


Figure 3.2 Time series plot of standardized residuals of GARCH(1,1). Data sample from 01 Dec 1977 to 31 Dec 2006.

Another relevant test of residuals involves examining their distribution. Because one important assumption of our GARCH(1,1) model is that the innovations z_t are conditionally normally distributed, it is necessary to check if this assumption holds with model's standardized residuals. Checking for normality can be done via inspecting histogram and Jarque-Bera test. Figure 3.3 indicates a leptokurtic distribution (with a excessive kurtosis of 18.35) and the Jarque-Bera test strongly rejects the null hypothesis of normality.

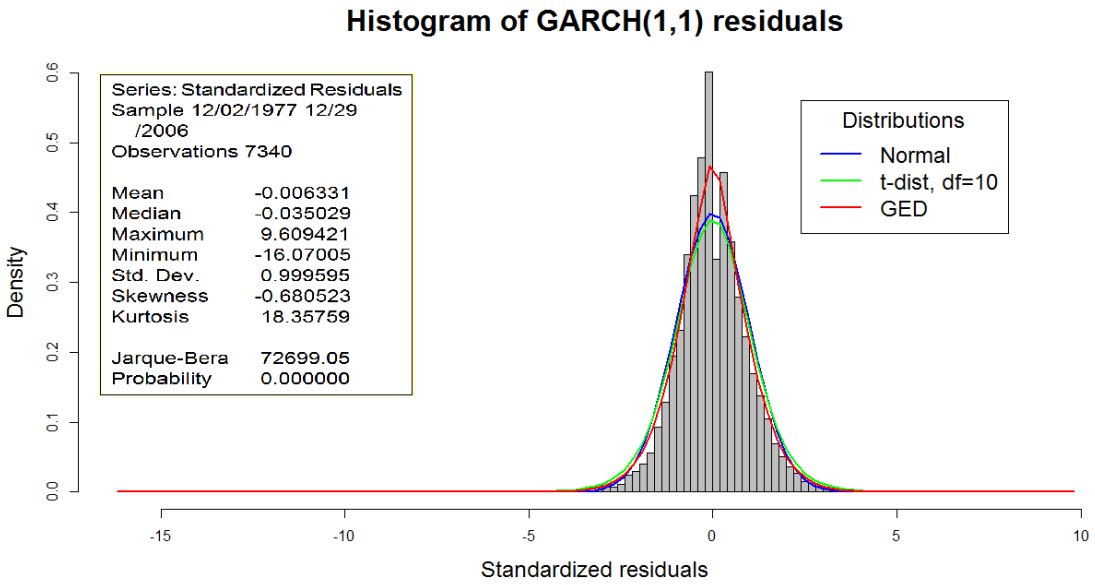


Figure 3.3 Histogram and descriptive statistics of standardized residuals for GARCH(1,1). Density curves from 3 distributions: normal, Student and GED with comparable mean and variance are superimposed.

Non-normality is also visible through the quantile-quantile plot of the empirical distribution of z_t against that of a Student distribution (with 10 degrees of freedom) and a Generalized error distribution (see Figure 3.4). Taken together, the histogram and Q-Q plot show that the deviation from normality seems to be significant. As we can see the empirical distribution of residuals may be better approximated by the t-distribution and GED than normal distribution. Interestingly, deviation from normality is driven by both extreme negative and positive returns shocks, which is consistent with the fact that GARCH(1,1) is unable to discriminate between the signs of shocks. Deviation from t-distribution and GED, on the other hand, seems to be mostly driven by large positive shocks.

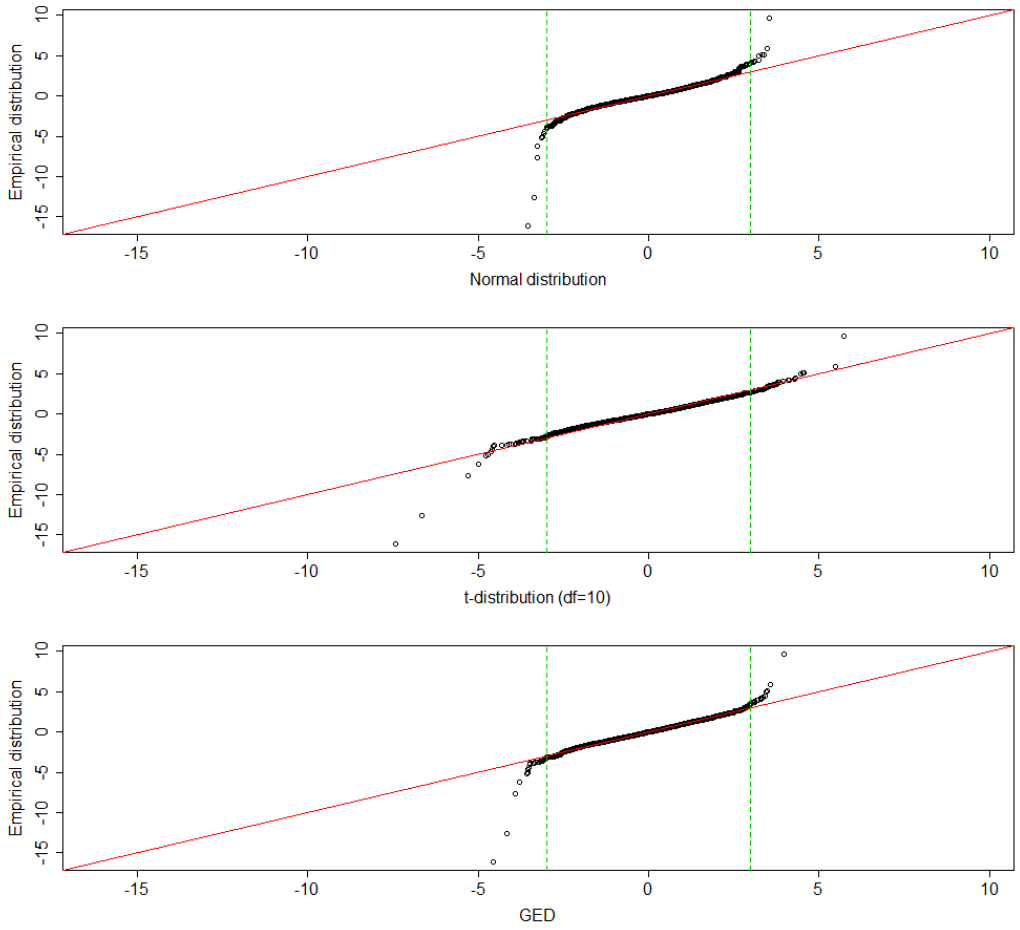


Figure 3.4 Q-Q plots of residuals' distribution against normal distribution (top), t-distribution (middle) and GED (bottom). Red line indicates identity line.

Given these insights, we proceed to adjust our original assumption of normality and re-estimate the GARCH models with Maximum likelihood functions corresponding to the Student t-distribution and Generalized errors distribution (note that our original estimates are robust to non-normality thanks to robust standard deviation proposed by Bollerslev and Wooldridge (1992)). Other conditions are preserved for re-estimation. Results are reported in Tables 3.2 and Table 3.3 for t-distribution and GED, respectively. Overall we observe improvement in terms of higher Log-likelihood values: for example, for GARCH(1,1) the maximum Log-likelihood increases from 19181.01 with Normal distribution to 19642.62 for t-distribution and 19587.99 for GED. Likewise the AIC decreases from -5.22 to -5.35 and -5.33. Estimated values of main parameters and corresponding significance do not change dramatically. However, this seemingly improved goodness-of-fit (with respect to empirical returns data) is not manifest in increased R-squared from regressing absolute returns against $\hat{\sigma}_t^2$. In particular, with GARCH(1,1) we obtain $R^2 = 0.1202$ for t-distribution and $R^2 = 0.1200$ for

GED (compared to $R^2 = 0.12042$ for normal distribution). However, all of the statistics associated with estimates for FIGARCH(1,1) model do not converge when the GED is adopted. The same observation is made with FIEGARCH(1,1) for both Student and GED distribution specifications, which is why we do not report the estimates of this model in these two cases. Nevertheless, throughout the various specifications of residuals' distribution, FIGARCH(1,1) exhibits a very high degree of fractional difference parameter (around 0.42).

Model estimates with Student t-distribution for residuals										
Data: Citigroup daily returns Range: 01/12/1977 - 31/12/2006 Number of observations: 7341										
Parameters	GARCH(1,1)		IGARCH(1,1)		EGARCH (1,1)		GJR-GARCH (1,1)		FIGARCH (1,1)	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
μ	0.000361* (0.000149)	0.015191	0.000358* (0.000155)	0.020812	0.000323* (0.000155)	0.037406	0.000312* (0.000152)	0.039472	0.000338* (0.000153)	0.0279
ω	0.000002 * (0.000001)	0.019990	0.000002* (0.000001)	0.011646	-0.080466 ** (0.03081)	0.009030	0.000002* (0.000001)	0.015692	0.000005** (0.000002)	0.0059
α	0.073575**** (0.016613)	0.000000	0.074869**** (0.016495)	0.000006			0.063886**** (0.014001)	0.000005	0.3645**** (0.0532)	0.0000
d									0.41726**** (0.03308)	0.0000
α^*							0.026275 (0.014973)	0.079293		
ϑ					-0.023312 (0.009434)	0.013469				
β	0.925424**** (0.016650)	0.0000	0.925131 N/A	N/A	0.990058**** (0.003824)	0.0000	0.921976**** (0.016527)	0.000000	0.64289**** (0.05381)	0.0000
γ					0.147611**** (0.029005)	0.0000				
Diagnostic information										
AIC	-5.3501		-5.3504		-5.3602		-5.3508		-5.3578	
BIC	-5.3454		-5.3466		-5.3546		-5.3451		-5.3522	
Log Likelihood	19642.62		19642.63		19680.6		19645.94		19671.98	
Sign bias	0.7066	0.4798	0.7039	0.4815	0.72644	0.46759	0.6252	0.5318	0.5243	0.6000
Neg. size bias	1.5279	0.1266	1.4890	0.1365	1.71230	0.08688	1.1377	0.2553	0.5052	0.6133
Pos. size bias	0.4140	0.6789	0.4423	0.6583	0.09818	0.92180	0.2846	0.7760	1.0637	0.2874
Joint Effect	2.5641	0.4638	2.4725	0.4803	2.94766	0.39976	1.3757	0.7112	1.5430	0.6723

Table 3.2 Summary of different GARCH family model estimates. Non-normality robust standard errors are reported in parentheses.
N/A indicates where no convergence was found for the corresponding statistic.
Statistical significance indicators: * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$). All models assume Student distribution of innovations.

Model estimates with GED for residuals										
Data: Citigroup daily returns Range: 01/12/1977 - 31/12/2006 Number of observations: 7341										
Parameters	GARCH(1,1)		IGARCH(1,1)		EGARCH (1,1)		GJR-GARCH (1,1)		FIGARCH (1,1)	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
μ	0.000002* (0.000010)	0.829188	0.000002 * (0.000009)	0.805652	0.000004 (0.000021)	0.857599	0.000001* (0.000006)	0.860649	0.000005 N/A	N/A
ω	0.000002 * (0.000001)	0.015597	0.000002* (0.000001)	0.012396	-0.094807 ** (0.031535)	0.002644	0.000003 * (0.000001)	0.014194	0.06563 N/A	N/A
α	0.074561**** (0.015852)	0.000003	0.075541**** (0.015860)	0.000002			0.066093 **** (0.013788)	0.000002	0.3780 N/A	N/A
d									0.4231 N/A	N/A
α^*							0.026198 (0.015497)	0.090931		
ϑ					-0.025049 (0.009647)	0.009417				
β	0.924438**** (0.015771)	0.0000	0.924459 N/A	N/A	0.988227**** (0.003917)	0.0000	0.919808**** (0.016504)	0.000000	0.65632 N/A	N/A
γ					0.154146**** (0.028132)	0.0000				
Diagnostic information										
AIC	-5.3352		-5.3355		-5.3448		-5.3357		-5.3423	
BIC	-5.3305		-5.3318		-5.3391		-5.3301		-5.3367	
Log Likelihood	19587.99		19588.01		19623.91		19590.85		19614.64	
Sign bias	0.6739	0.5004	0.6726	0.5012	0.629	0.5294	0.5924	0.5536	0.4967	0.6193
Neg. size bias	1.5415	0.1232	1.5047	0.1324	1.608	0.1078	1.1376	0.2553	0.4756	0.6343
Pos. size bias	0.4012	0.6883	0.4290	0.6679	0.119	0.9052	0.2823	0.7777	1.0911	0.2751
Joint Effect	2.6086	0.4560	2.5208	0.4715	2.602	0.4571	1.3766	0.7110	1.6149	0.6560

Table 3.3 Summary of different GARCH family model estimates. Non-normality robust standard errors are reported in parentheses.
N/A indicates where no convergence was found for the corresponding statistic.
Statistical significance indicators: * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$). All models assume Generalized error distribution of innovations.

3.4 Out-of-sample GARCH forecast performance

The primary value of a time series model lies in its ability to provide reliable approximations of the modelled variable, both in-sample (where data is used to estimate model parameters) and out-of-sample (where the model is updated with new data and produce forecasts). Here we illustrate how high-frequency data is used to construct a 'realized' volatility measure which may provide a better approximation to the latent volatility variable. For comparative purposes, in the first part of the subsequent analysis we adopt the approach of Andersen and Bollerslev (1998) and Hansen and Lunde (2005). In line with these authors, we find some evidence for the credibility of GARCH(1,1) model. Subsequently, we show that when using these GARCH estimates as an indicator, our wavelet based multi-scale decomposition may potentially provide an improved approximation to the true volatility factor. It should be noted that, throughout the section, we only focus on 'ex-post' one-period-ahead forecasts when evaluating model performance out-of-sample, meaning we use known returns data to update our estimated GARCH model and get the forecasts of volatility (a technique known as "rolling forecast"). Although these forecasts have less practical value than the 'ex-ante' forecasts (e.g. rather than using known returns to get volatility forecasts, we may first forecast returns themselves), this is a standard approach which provides an objective and direct assessment of model performance.

3.4.1 Daily volatility vs. realized volatility

3.4.1.1 Study design

To begin we use a more general notation of returns proposed by Andersen and Bollerslev (1998), which incorporates the sampling frequency m:

$$r_{(m),t} = p_t - p_{t-1/m} \text{ where } \; t = 0, 1/m, 2/m, \ldots$$

in which p_t denotes the logarithmic price. With the sampling rate m we define $\{r_{(m),t}\}$ as the discretely observed time series of continuously compounded returns with m observations per day. The instantaneous returns process is then defined via $r_t \equiv r_{(\infty),t} \equiv dp_t$. When $m = 1$ we have the daily returns, which represents the standard frequency in many studies. Following Andersen and Bollerslev (1998) we model the (1-period) daily volatility using the generalized GARCH(1,1)

specification:

$$\begin{aligned}
&\sigma^2_{(m),t} = \omega_{(m)} + \alpha_{(m)}.r^2_{(m),t-1/m} + \beta_{(m)}.\sigma^2_{(m),t-1/m} \\
&r_{(m),t} = \sigma_{(m),t}.z_{(m),t} \\
&\text{where } \ \omega_{(m)} > 0, \alpha_{(m)} \geq 0, \beta_{(m)} \geq 0 \\
&z_{(m),t} \sim \text{i.i.d } \ \text{N(0,1)} \\
&\text{where } m = 1
\end{aligned} \tag{3.4.1}$$

Theoretically, if we have a correctly specified model, our forecast must equal the true return variance, i.e.:

$$E(r^2_{t+1}) = E(|r_{t+1}|^2) = E(u^2_{t+1}) \equiv \hat{\sigma}^2_{t+1}$$

This motivates a direct way to evaluate GARCH forecast performance, that is, to use some type of MSE (mean squared errors)-based metrics, that is, to test for the null hypothesis: $E(\hat{\sigma}^2_{t+1} - \text{Var}[r_{t+1}]) = 0$. To do this the most popular approach is analogous to Mincer and Zarnowitz (1969)'s method to evaluate conditional mean forecasts (but in this context we need to evaluate conditional variance forecasts instead). In particular, we simply need to regress some ex-post realized volatility measurement on the estimated values obtained from GARCH(1,1) as follows:

$$r^2_{(m),t+1/m} = a_{(m)} + b_{(m)} \ . \ \hat{\sigma}^2_{(m),t+1/m} + u_{(m),t+1/m} \tag{3.4.2}$$

Now the next step is to choose an appropriate ex-post volatility proxy.

Andersen and Bollerslev (1998) criticized the use of squared returns as a volatility proxy. They show that using this quantity results in a 'systematically' poor forecast performance, because daily squared return is an unbiased, but very noisy estimator of the true latent volatility. This is evident in a 'disappointingly' low R-squared obtained when regressing this proxy on the forecasts, as commonly observed in numerous contemporary papers. In hope of reducing the degree of misspecification associated with using squared returns by a power of 2, we opt for using absolute returns. This is also consistent with the view of volatility as standard deviation rather than variance. In addition, Taylor (1986)'s findings suggest that better forecasts of volatility were provided by models using absolute returns compared to squared returns. This might be a result of absolute returns displaying higher degree of long-memory than squared returns, which motivates direct modelling of volatility from absolute returns (See e.g. Ding and Granger (1996) and Vuorenmaa (2005)). Furthermore, as pointed out by Vuorenmaa (2005), since the logarithmic squared value of a close-to-zero return would be negative, it is less favourable to be used as volatility proxy compared to an absolute return,

which is more robust to this kind of 'inlier' problem. The above arguments have been largely verified with stock data, at the very least.

3.4.1.2 Implementation

We start with estimating the GARCH(1,1) model (specified in Equation 3.4.1, using in-sample data based on daily returns of Citigroup, Inc. from 01 Dec 1977 to 31 Dec 2006. Next, we update the estimated model with out-of-sample daily returns from 03 Jan 2007 to 31 Dec 2007 and get the rolling forecast of volatility (denoted as $\hat{\sigma}_{(m),t+1/m}$). Consequently, our one-step ahead GARCH(1,1) volatility forecasts are computed for the period of 250 working days of the year 2007 (excluding weekends and holidays for their insignificant trading activities). We then proceed to replace the fitted/forecast variance in equation 3.4.2 with these values. In other words we modify Andersen and Bollerslev (1998)'s approach by regressing absolute returns on forecast standard deviation instead of regressing squared returns on forecast variances [20]:

$$|r_{(m),t+1/m}| = a_{(m)} + b_{(m)} \; . \; \hat{\sigma}_{(m),t+1/m} + \epsilon_{(m),t+1/m} \tag{3.4.3}$$

Substituting absolute daily returns ($m = 1$) into this generalization we have:

$$|r_{(1),t+1}| = a_{(1)} + b_{(1)} \; . \; \hat{\sigma}_{(1),t+1} + \epsilon_{(1),t+1} \tag{3.4.4}$$

The R-squared value of Equation (3.4.4) ($R^2_{(1)}$) can be viewed as a simple indicator of how well our estimates of volatility can explain the variability in the ex-post returns which, in theory, can not be observed. The problem remains, as discussed earlier, that traditional daily proxies of volatility (e.g. squared residual returns, squared returns, absolute returns) may not be adequate. In particular, using daily absolute returns as volatility proxy (or the dependent variable in Equation 3.4.4) suggests that our GARCH(1,1) model only explains 13.88% ($R^2_{(1)} = 0.1388$) of daily variability of returns. When using squared returns we obtain a decreased value of $R^2_{(1)} = 0.0997$, but this is still higher than the values reported in Andersen and Bollerslev (1998)'s exchange rate study, i.e. $R^2_{(1)} = 0.047$ and $R^2_{(1)} = 0.026$ for the Deutsche Mark/U.S. Dollar (DM-$) and Japanese Yen/U.S. Dollar (¥-$) rates, respectively.

All things considered, these results suggest that absolute returns may provide

[20] The former specification is also adopted by Jorion (1995). A viable alternative would be to compare volatility forecasts with ex-post returns shocks (or residual returns) out-of-sample. But as Tsay (2001) argued, using returns shocks as a measure of realized volatility may not be a good idea. A single realization of the random variable u^2_{t+1} is not adequate to provide an accurate estimate of the movement of day-by-day volatility (or $\text{Var}_{t+1}(u_t)$).

a more appropriate volatility proxy than squared returns. However, regardless of the proxies used, the low $R^2_{(1)}$ and a poor out-of-sample fit are to be expected, as both proxies are noisy estimators of true volatility process. In defence of the GARCH model, Andersen and Bollerslev (1998) assert that the use of these noisy proxies distort our perception of GARCH forecasts performance, because "*[...] Rational financial decision making hinges on the anticipated future volatility and not the subsequent realized squared returns. Under the null hypothesis that the estimated GARCH(1,1) model constitutes the correct specification, the true return variance is, by definition, identical to the GARCH volatility forecast.*" (p.890).

Realized volatility proxy constructed from intraday returns The above arguments imply that to obtain meaningful forecast evaluation we first need to acquire a better volatility estimate. According to the theory of quadratic variation (See Karatzas and Shreve (1991)), in principle, the cumulative squared intraday returns provide better approximation to ex-post volatility at higher sampling rate:

$$\lim_{m\to\infty}\Big(\int_0^1 \sigma^2_{t+\tau}d\tau \ - \sum_{j=1}^{m} r^2_{(m),t+j/m}\Big) = 0$$

in which $\int_0^1 \sigma^2_{t+\tau}d\tau$ denotes the continuous time volatility measure and $r^2_{(m),t+j/m}$ is the length j/m period squared return. Motivated by this proposition, Hansen and Lunde (2005) show that when intraday returns are uncorrelated we have an unbiased estimator of σ^2_t simply by summing intraday squared returns:

$$\begin{aligned} \hat{\sigma}^2_{(m),t} &\equiv \sum_{j=1}^{m} r^2_{(m),t+j/m} \\ E\left[\hat{\sigma}^2_{(m),t}\right] &= E(r^2_t) = E\left[\hat{\sigma}^2_{(1),t}\right] \end{aligned} \tag{3.4.5}$$

Ideally we would choose m close to infinity. In practice, irregularly spaced price realizations would render computing 'continuous' measurement at an infinite sampling rate infeasible. Furthermore, various microstructure effects may result in negative serial correlation among returns (Dacorogna et al. (2001)). Many researchers suggest 5 minutes as the optimal sampling interval. Therefore we opt for using a realized volatility measures computed from intraday returns recorded every 5 minutes for the year 2007 (the detailed formulation and properties of this series will be explored in the next Chapter). These returns are calculated from a value weighted average of tick-by-tick prices provided by Sirca database. The first and last tick of the day are omitted in this calculation, as their corresponding

trading volumes are much higher than those of the rest of the ticks. These 'outliers' are the result of catching up with cumulative overnight information.

For assets that are not traded continuously throughout the day, we can only construct a 'partial' realized volatility measure using the observed intraday data, that is:

$$\hat{\sigma}^2_{(m,p),t} \equiv \sum_{j=1}^{p} r^2_{(m),t+j/m}$$

Naturally, this partial measure underestimates the total daily variation of returns. However, Hansen and Lunde (2005) point out that we can scale our partial measure with the inverse of the fraction of the day when we obtain intraday returns, i.e. $\frac{p}{m}$, to correct for this bias. Since the New York Stock Exchange (NYSE)'s standard trading hour is from 9:30 E.S.T to 16:00 E.S.T (6.5 hours) we have $p = 78$ observations (or intervals) per day (instead of a full day of $m = 288$ in most studies concerning 24-hour foreign exchanges trading). This gives us a total of 19500 five-minute observations for 250 trading days. Because the sum of squared 5-minute returns in our sample only represents roughly 6 hours of trading, we compute daily realized volatility by multiplying (or rescaling) this sum by 4 (using $\frac{p}{m} = \frac{78}{288} = 3.69$ would result in little difference) [21].

$$RV_{5min,t} = \sqrt{\left(\sum_{j=1}^{78} r^2_{(78),t+j/78}\right) \times 4} \quad (t = 1, \ldots, 250)$$

so that the return-volatility regression (specified by Equation 3.4.4) is modified to be:

$$|RV_{5min,t+1}| = \alpha_{(1)} + \beta_{(1)} \; . \; \hat{\sigma}_{(1),t+1} + u_{(1),t+1} \tag{3.4.6}$$

3.4.1.3 Results

Our measure of ex-post realized volatility improves the goodness-of-fit of GARCH estimates significantly. In particular, the coefficient of multiple determination increases from $R^2 = 0.1388$ to $R^2 = 0.5165$. Summary for these forecast performance comparisons are reported in Table (3.4). With our simple realized measurements we can obtain a great improvement in ex-post forecast performance, which can be seen in Figure 3.5. To compare our approach to that of Andersen and Bollerslev (1998), we substitute squared RV_{5min} and $\hat{\sigma}^2_t$ for re-

[21] Hansen and Lunde (2005) mentioned a simpler method to account for the bias associated with the use of only a proportion of 24 hours: adding the squared returns computed from closing price and next day's opening price. Although the addition accounts for the non-trading 3/4 portion of the day, it is almost as noisy as daily squared returns, since the added component disregards unobserved intraday fluctuations during the non-trading session.

gression and obtain $R^2 = 0.4341$. This value is comparable to that documented by these authors, who report R-squared values of 0.479 and 0.392 for DM-$ and ¥-$ series, respectively.

Panel A: Regression with absolute returns

$$|r_{(1),t+1}| = a_{(1)} + b_{(1)} \cdot \hat{\sigma}_{(1),t+1} + \epsilon_{(1),t+1}$$

	Estimate	Std. Error	p-value	R^2
a	0.000140	0.002115	0.947	0.138885
b	0.711384***	0.112255	0.0000	

Joint test for model forecast performance ($H_0 : a = 0, b = 1$)

Residual d.f	RSS	Sum of square	F-stat	p-value
250	0.0449			
248	0.0379	0.00707	23.229***	0.0000

Panel B: Regression with realized volatility measures

$$|RV_{5min,t+1}| = \alpha_{(1)} + \beta_{(1)} \cdot \hat{\sigma}_{(1),t+1} + u_{(1),t+1}$$

	Estimate	Std. Error	p-value	R^2
α	-0.000359***	0.001624	0.0000	0.5165
β	1.41***	0.086312	0.0000	

Joint test for model forecast performance ($H_0 : \alpha = 0, \beta = 1$)

Residual d.f	RSS	Sum of square	F-stat	p-value
250	0.035502			
248	0.02222	0.013282	74.119***	0.0000

*** indicates p-value ≤ 0.0001.

Table 3.4 GARCH(1,1) (ex-post) out-of-sample forecasts performance regression results.

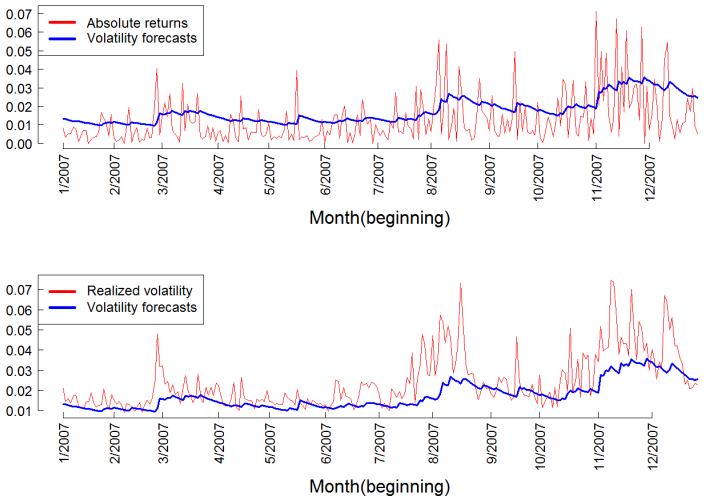


Figure 3.5 Comparing one-step ahead daily forecasts from GARCH(1,1) model ($\hat{\sigma}_{(1),t+1}$) with corresponding absolute returns (top) and realized volatility measure computed from 5-minute returns (bottom). Data sample ranges from 01 Jan 2007 to 31 Dec 2007.

It is worth noting that while the slope coefficients reported in Table 3.4 are highly significant, the intercept coefficients are insignificant and have a value close to zero. Given a correctly specified model, we would observe $a = 0, b = 1$ where a and b are the coefficients from the above regressions. In practice, estimation error would introduce a downward bias in the estimate of b (See e.g. Chow (1983) and Christoffersen (1998)). Therefore by testing the null hypothesis ($H_0 : a = 0, b = 1$) we have a means to determine whether our model can provide good forecasts (Mincer and Zarnowitz (1969)). F-test outputs reported in table 3.4 implies rejection of the null hypothesis for both regressions. This shows our estimates do not track the RV measures as well as we would like, even though in theory they are superior to the daily measures. This can possibly be explained by an argument from Awartani and Corradi (2005), that while RV may be a better unbiased estimator than daily (squared) returns, it is only consistent when the underlying returns follow a continuous semi-martingale, which is not true for discrete GARCH process. In other words, RV does not outperform daily measure significantly in terms of representing the true latent volatility in the context of a discrete time data generating process.

To some extent we have accounted for the misspecified ex-post volatility measure associated with daily returns. Now we can utilize the 6 loss functions suggested by Hansen and Lunde (2005) as measures for performance of our models out-of-sample, all of which are modifications of the MSE whose lower value indicates better forecasts. In particular, given $n = 250$ we would have:

$$\begin{aligned} MSE_1 &= n^{-1}\sum_{t=1}^{n}(\hat{\sigma}_t - h_t)^2 \;\; ; \;\; MSE_2 = n^{-1}\sum_{t=1}^{n}(\hat{\sigma}_t^2 - h_t^2)^2 \\ PSE &= n^{-1}\sum_{t=1}^{n}(\hat{\sigma}_t^2 - h_t^2)^2 h_t^{-4} \;\; ; \;\; R2LOG = n^{-1}\sum_{t=1}^{n}\left[\log(\hat{\sigma}_t^2 h_t^{-2})\right]^2 \\ MAD_1 &= n^{-1}\sum_{t=1}^{n}|\hat{\sigma}_t - h_t| \;\; ; \;\; MAD_2 = n^{-1}\sum_{t=1}^{n}|\hat{\sigma}_t^2 - h_t^2| \end{aligned} \tag{3.4.7}$$

where σ_t is the estimated volatility and h_t indicates our ex-post measures (RV_{5min}). Results are reported in Table 3.5 , which shows little difference between the models, although there is slightly better performance of GJR-GARCH(1,1). Overall, all 4 models examined provide relatively good out-of-sample forecasts. Again the results need cautious consideration regarding the limit of data range.

It can be observed that, though we do not obtain relatively high values of in-sample R-squared as documented by other researchers, when considering the 6 metrics previously described it seems that models accounting for asymmetry do

have slightly better performance, both in and out-of-sample.

	Forecast performance metrics					
Model	MSE_1	MSE_2	PSE	$R2LOG$	MAD_1	MAD_2
GARCH(1,1)	0.00014200	8.20×10^{-7}	0.259646	0.720293	0.007773	0.000454
IGARCH(1,1)	0.00014139	8.17×10^{-7}	0.260498	0.719489	0.00776	0.000453
EGARCH(1,1)	0.00013793	8.07×10^{-7}	0.255495	0.690906	0.00769	0.000450
GJR-GARCH(1,1)	0.00013171	7.74×10^{-7}	0.250268	0.668722	0.00750	0.000439

Table 3.5 Comparing out-of-sample forecasts performance for different GARCH models, using 6 metrics suggested by Hansen and Lunde (2005).

In conclusion, the empirical evidence presented in this subsection indicates that to some extent, the volatility process is predictable and/or can be effectively modelled by GARCH models. In addition, this simple example illustrates the important role of intraday data in constructing improved ex-post volatility measurements. The results (especially for the joint test of $H_0 : a = 0, b = 1$), although being more or less consistent with previous studies, remain somewhat inconclusive, considering the (discrete) out-of-sample data range is relatively small compared to the large in-sample data range.

3.4.2 New insights from wavelet-based decompositions

In Appendix A we discuss the methodology of wavelet-based Multi-Resolution Analysis. Here we applied the MRA (with an LA(8) MODWT) to decompose the RV time series at different levels and retain the corresponding smooth series. Note that the smooth series, denoted as S_j, where $j = 1, \ldots, J$ is the scale/level indicator, can be viewed as the underlying trend, or the sample average of the original signal after extracting all variations associated with scale smaller than λ_j (or with frequencies higher than $\frac{1}{2^{j+1}}$). As shown in Figure 3.6, using the increasingly smooth series as substitutions for our RV measure in regression 3.4.6, we are able to obtain higher R-squared values. However, as soon as the smoothing 'depth' reaches level 6, the goodness-of-fit starts decreasing (we can only decompose the RV series up to level 7, the highest our data range allows). The highest R^2 value that can be obtained is 0.7675 at level 5 (about 50% larger than the R^2 corresponding to RV regression). Apparently, the more high-frequency fluctuations filtered out of RV, the better our GARCH estimates perform. As such, we would recommend this approach as a follow-up to extend Andersen and Bollerslev (1998) and Hansen and Lunde (2005)'s analyses: that is, we can use a smooth component as a better ex-post volatility estimate, which has better compatibility to GARCH(1,1) forecasts (compared to the original RV measure). This may fur-

ther supports these authors' conclusion regarding the credibility of the GARCH model.

An important observation from this case study is that the degree of goodness-of-fit peaks at level 5, when all fluctuations (represented by the detail series) associated with frequency band ranging from $\frac{1}{2^6}$ to $\frac{1}{2}$ are filtered out, leaving only activities associated with frequencies lower than $\frac{1}{2^6}$. As the original RV series is a daily series, this means the level 5 smooth series are 'freed' of activities associated with a period length between $2^1 = 2$ days and $2^6 = 64$ days (roughly enclosing the time frame of 2 months). However, when the activities associated with period length from 64 days to 128 days are excluded, R-squared reverts back to approximately the level obtained from regressing the original RV (around 0.5). In sum, in terms of constructing a good ex-post volatility estimate, the inclusion of fluctuations at frequencies lower than $\frac{1}{128}$ is inadequate while the inclusion of fluctuations at frequencies higher than $\frac{1}{64}$ is redundant.

All things considered, this fact implies that the collective activities associated with a trading horizon longer than 2 months (represented by level 5 smooth series) constitute the 'core' component of the true latent volatility process. Granted, this argument can only be valid provided the credibility of two things: (i) this process can be reasonably proxied by our RV measure and (ii) the GARCH model is correctly specified.

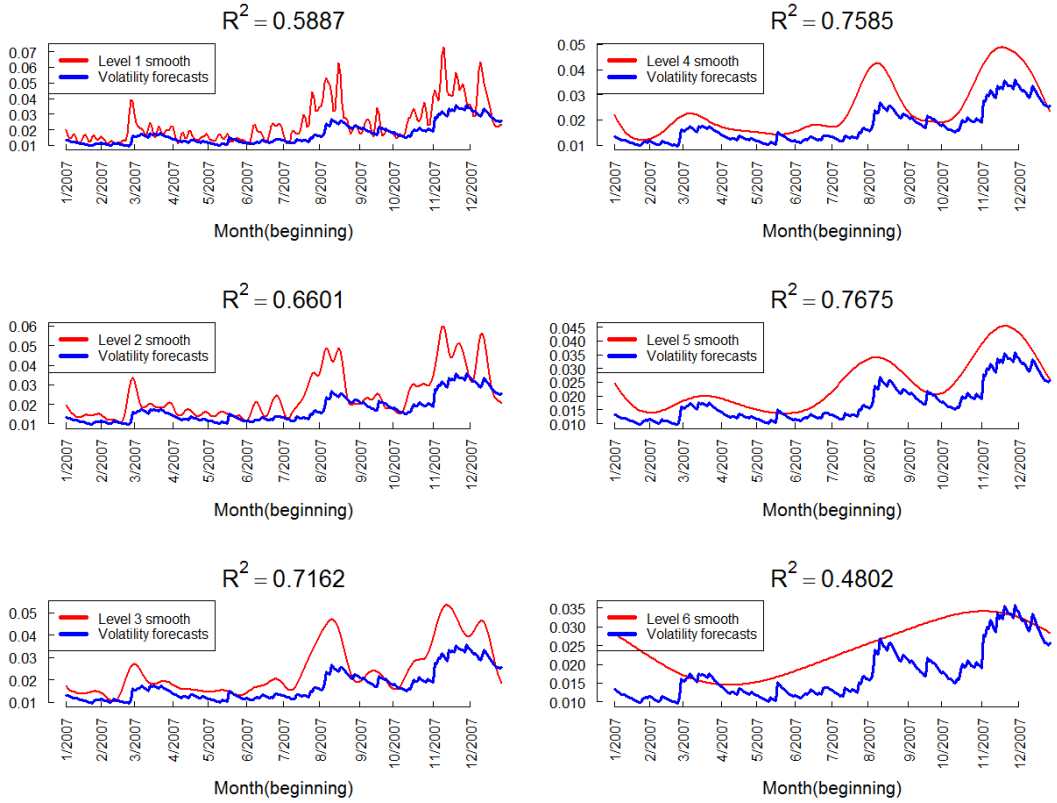


Figure 3.6 Comparing one-step ahead forecasts from GARCH(1,1) model with smooth series obtained from an MRA of realized volatility series of Citigroup. Decomposition depth is set to 6 levels.

Remark Here we re-emphasize the twofold implication of our major finding from this section:

- Despite their simplicity, GARCH models can provide decent forecasts when we use RV as proxy for the true latent volatility factor. Typically, the forecasts account for close to 50% of variability in ex-post volatility, which is in line with Andersen and Bollerslev (1998) and Hansen and Lunde (2005).
- If we believe in the indicative value of GARCH estimate, then the smoothed RV corresponding to a trading period longer than 2 months may be a close approximation to the true volatility process.

3.5 Evidence of leverage effect

As discussed in subsection C.3, the negative correlation between unexpected returns and future volatility is widely documented and has motivated the vast literature revolving around leverage effect hypothesis. Intuitively, stockholders

recognize higher risk following a fall in stock price and increase in leverage ratio. In our study, evidence supporting this hypothesis can be obtained with the application of GJR-GARCH and EGARCH models. Specifically, the larger impact of negative returns shocks compared to positive ones is visible in these models' news impact curves:

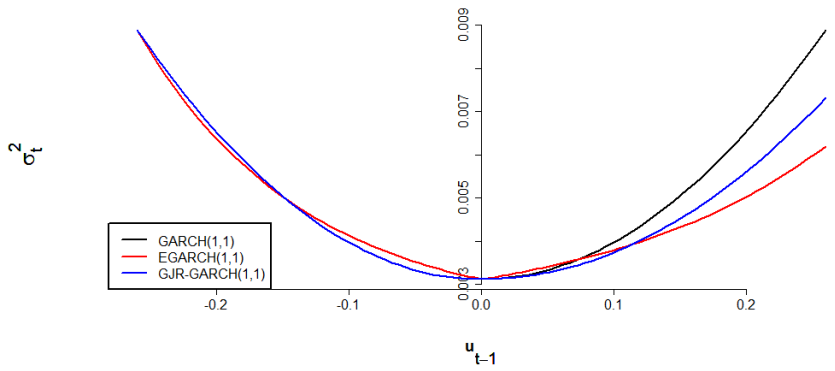


Figure 3.7 News impact curves of different asymmetric GARCH models for daily Citigroup volatility. In-sample data range from 01 Dec 1977 to 31 Dec 2006.

We may see that the difference between the two 'wings' of the asymmetric curves is considerable [22]. Specifically, with the EGARCH model, a 20% unexpected increase in return leads to an estimate of next period's conditional variance (σ_t^2) of 0.00187 while a return decrease of the same magnitude results in a σ_t^2 estimate of 0.00293. These values correspond to volatility estimates of 4.32% and 5.41%, respectively (note that volatility is measured by standard deviation, or σ_t). The gap is wider with higher, or more extreme changes in returns. Similarly, for the GJR-GARCH model, the value of σ_t^2 are 0.0066 following a positive 20% shock and 0.0081 following a negative one. These values correspond to volatility estimates of 8.12% and 9%, respectively. These analyses are illustrated in Figure 3.8.

In addition, consistent with the findings in Engle and Ng (1993) and Nelson (1991), we documented significant negative value of ϑ and positive value of γ for our EGARCH model, indicating larger impact of negative returns shocks (for further discussions, see section C.3, Appendix C). Likewise, positive value of α^* in GJR-GARCH model implies similar conclusion. Interestingly, the statistical significance of our estimates of asymmetric-related parameters increases as we adjust the assumption of residuals' distribution. In particular, with normal distribution specification we obtain significance at the 10% level for ϑ and insignificant α^*. Estimates of ϑ are significant at the 5% level (for t-distribution) and 1% level

[22] Although it is not as striking as that of the S&P500 series (see Figure C.2, Appendix C).

(for GED). Estimates of α^* are significant at 10% for both t-distribution and GED. Estimates of γ remains highly significant across all distribution assumptions. These estimates are reported in Tables 3.2 and 3.3.

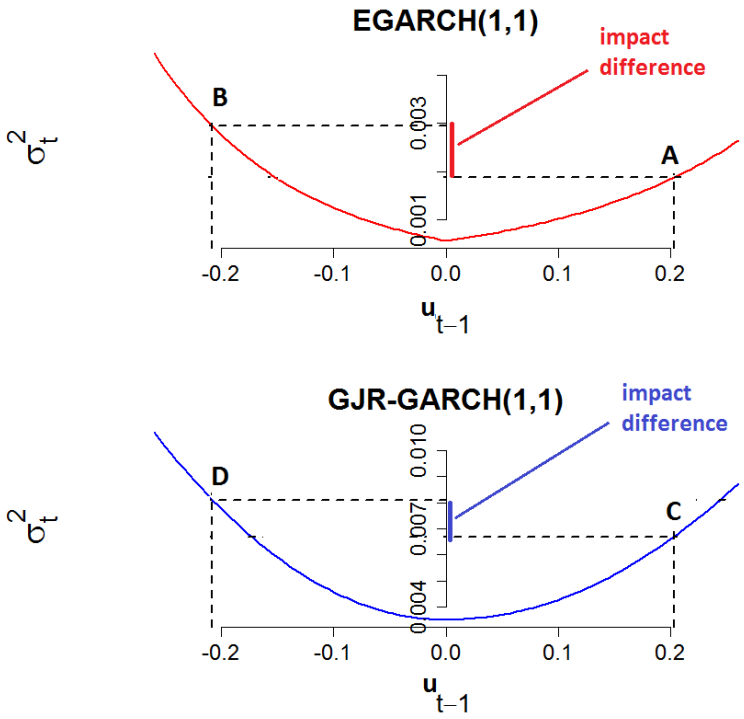


Figure 3.8 Different impact of good news (points A and C) and bad news (points B and D) implied by EGARCH and GJR-GARCH models. In-sample data range from 01 Dec 1977 to 31 Dec 2006.

Another, more general way to examine the asymmetric effect of return shocks is to plot the cross correlation function (CCF) [23] between returns and volatility, proxied by daily absolute returns and our realized measure. In Figure 3.9 we show the CCF using both in-sample and out-of-sample data. As we can see, up to around lag 40, most of the coefficients for positive lags are negative and are much higher than that for negative lags. This implies long-lasting (negative) impact of past returns movements on present volatility. Although there are not enough significant coefficients to validate this long term effect, lag 1 coefficient is -0.05 which is significant at the 5% level and helps support the results of EGARCH and GJR-GARCH models. Similar observations can be derived from the two out-of-sample CCFs, where maximum correlation coefficients are observed at lag 16 (0.179) and lag 15 (0.250) when using daily absolute returns and realized volatil-

[23] The CCF is computed as a conventional Pearson product-moment correlation, with 95% confidence interval is specified as $[\pm 1.96/\sqrt{n}]$ as per standard theory (see e.g. Venables and Ripley (2002)).

ity, respectively. All in all, it seems that our empirical results strongly support the leverage effect hypothesis at the daily level.

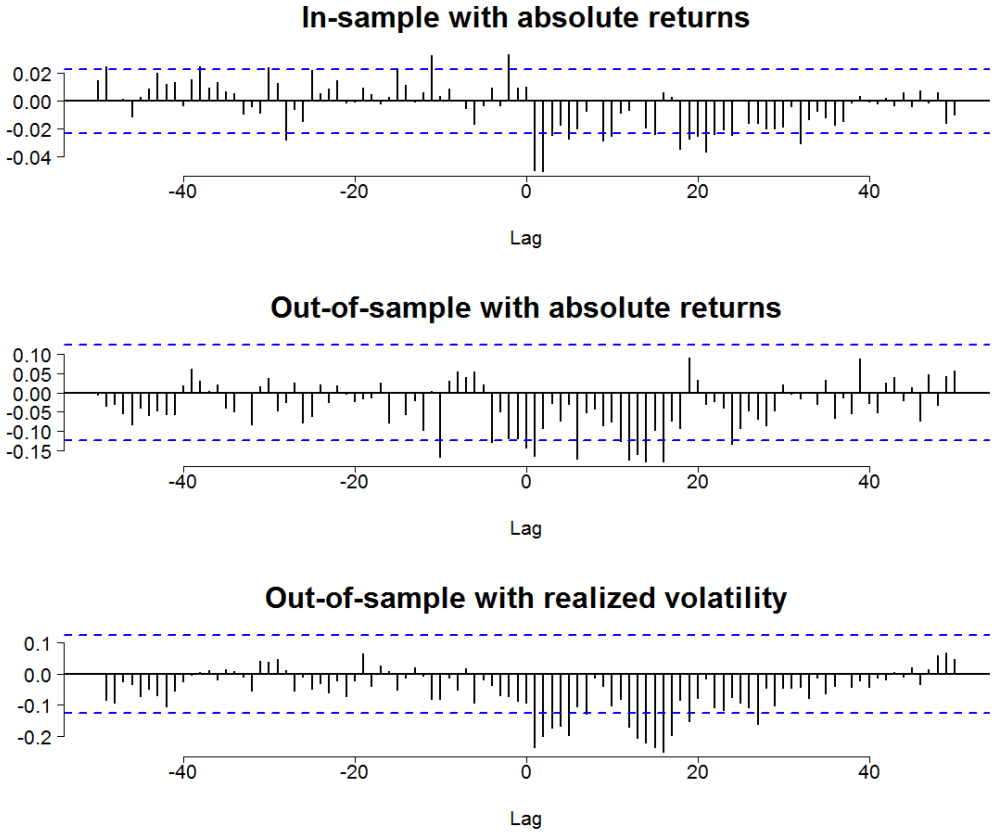


Figure 3.9 Cross correlation between daily returns and volatility. In-sample data range from 01 Dec 1977 to 31 Dec 2006. Out-of-sample data range from 01 Jan 2007 to 31 Dec 2007. Blue dashed lines indicate the 95% confidence intervals.

It can be seen that the leverage effects seems to be much more profound when considering the out-of-sample data. This may be explained by the highly volatile period of Citigroup's stock returns during the beginning of the GFC. In this turbulent time when market was heavily affected by news speculations, large returns decrease could have had a longer effect than in 'normal' times represented by the 'in-sample' period. Additionally, for the same period the realized measure, which is considered a better volatility proxy, reveals a stronger asymmetric effect compared to the original daily measure.

Chapter 4

Empirical studies using wavelet method

Preamble

In this Chapter we gather together the materials explored in the earlier discussions and investigate them in the context of a particular company, Citigroup, Inc. We also extend our study to various topics of relevance to volatility structure, which have not yet been covered in previous chapters.

This chapter assumes the following outline: Section 4.1 provides a brief description of the general data treatment techniques we adopt to cope with missing data. From this general discussion we conclude that it is reasonable to follow conventional practice and concatenate our data. In subsection 4.1.1 we examine the descriptive statistics of daily returns and volatility and propose a winsorizing procedure to mitigate the impact of outliers. In addition prominent stylized facts associated with the autocorrelation structure and returns distribution are studied. Next, subsection 4.1.2 examines whether similar stylized facts are exhibited by the intraday sample, which is firstly filtered to account for possible daily seasonality that may affect further analyses. In short, the first section provides preliminary analyses leading up to the main study.

Section 4.2 presents the primary results of the application of two main methodologies discussed in previous chapters: the GARCH model of volatility and the range of long-memory estimating techniques. Our major contribution to these approaches is the incorporation of the perspective of multi-scale/multi-frequencies via various wavelet-based techniques. This helps unveil some important intuition regarding the underlying process that governs return-volatility dynamics and dependence structure. Consequently, we are able to reconfirm the conclusion of Chapter 3 regarding the high goodness-of-fit of our GARCH models at lower frequencies, thus supporting the importance of trading activities at longer horizons. In addition, our estimated long-memory parameter suggests a certain predictability of volatility that varies across time scales. Furthermore, we proceed to emphasize the link between some of the most popular trading strategies based

on technical analysis and the long-run dependence behaviour of stock returns.

Throughout this Chapter, daily returns are winsorized to account for outliers and intraday returns are adjusted for microstructure effects. In addition, volatility is defined as the absolute value of returns, unless explicitly stated otherwise. As discussed in Chapter 3, when treating the long-run conditional returns as zero, there is no difference between returns and residual returns in the GARCH models.

4.1 Preliminary examination of data

Our main focus in this section would be to provide a preliminary investigation on Citigroup's stock prices and returns over the last 30 years or so. More emphasis shall be placed on the most recent decade when the firm experienced its hardest time during the GFC and subsequently recovered from the brink of bankruptcy. A brief introduction to the background of Citigroup is provided in Appendix D.

4.1.1 Daily data

4.1.1.1 Concatenation vs. linear interpolation of returns

Assume each day's return is an i.i.d random variable with zero mean and variance V, in accordance with well-established stylized facts (see for example, Taylor (2005)), and with V varying relatively slowly. Denote r_i as return of day i (i indicates day-of-the-week, e.g. r_1 is the one-day return computed using Monday closing price and Tuesday closing price). Since we generally do not have any data over weekends, we have to find a way to overcome this issue. As mentioned earlier, the usual remedy is to use concatenated returns and an obvious alternative is to linearly interpolate over missing data.

The concatenation method computes Friday's returns using Friday's and next Monday's closing price, resulting in a weekly sequence of returns $R_{con} = \{r_1, r_2, r_3, r_4, r_5'\}$ in which $r_5' = r_5 + r_6 + r_7$ (r_6 and r_7 are 'supposed' returns over Saturday and Sunday). This approximation is reasonable since returns are measured as the differenced series of price logarithms, so that returns are approximately additive. Furthermore, with the assumption of zero mean return, the variance of the return variable is the expected value of squared returns. We know that volatility is essentially the standard deviation of returns. In the case of a single day, our best estimate of this value is the absolute value (or squared value) of the return on that day. The same approach is applied for holidays. In an ideal market returns are expected to be uncorrelated, leading to the fact that variance of a sum of returns equal the additive sum of each

day's return variance. That is, for a (concatenated) five-day week we have: Var $(r_1+r_2+r_3+r_4+r_5') = E[r_1^2+r_2^2+r_3^2+r_4^2+(r_5')^2] = 7V$. This method is the simplest remedy: we just use whatever data available to compute returns over periods containing missing data, which are mostly associated with weekends and holidays. An alternative is linear interpolation: Instead of using r_5' as a representative of three consecutive daily returns, we split it into three equal values and allocate these to each day from Friday to Sunday. The resulting returns sequence is thus $R_{int} = \{r_1, r_2, r_3, r_4, \frac{r_5'}{3}, \frac{r_5'}{3}, \frac{r_5'}{3}\}$ [24]. Neither of these methods seem to account convincingly for overnight returns (Cheng, Roberts, and Wu (2013)). As a matter of fact there is a vast ongoing literature debating the best practice to deal with overnight returns, which is beyond the scope of this paper. We simply follow the conventional approach of concatenation.

4.1.1.2 Data description

In this section we shall explore some preliminary analyses of daily returns and volatility, with emphases on the so-called 'stylized facts', specifically the heavy-tailed distribution and the autocorrelation structure. We collect daily closing prices of Citigroup between 03 Jan 1977 and 31 Jul 2013 from `http://finance.yahoo.com`. To get a first impression of the data, following the work of Cheng et al. (2013), we simply concatenate the price series over the weekends and holidays. This means we effectively ignore any missing data in our sample of 30 years daily data, obtaining a total of 9228 daily returns. Figure 4.1 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 concatenating method). Some most 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 Enron scandal (2002) as well as the recent GFC (about right 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 one-period continuously compounded returns is established as in Chapter 3, where P_t is the adjusted closing price at time t, i.e. $r_t = \ln P_t - \ln P_{t-1}$.

[24] By construction, the variance of individual concatenated return should be approximated by $7V/5$ while that of interpolated return is $5V/7$.

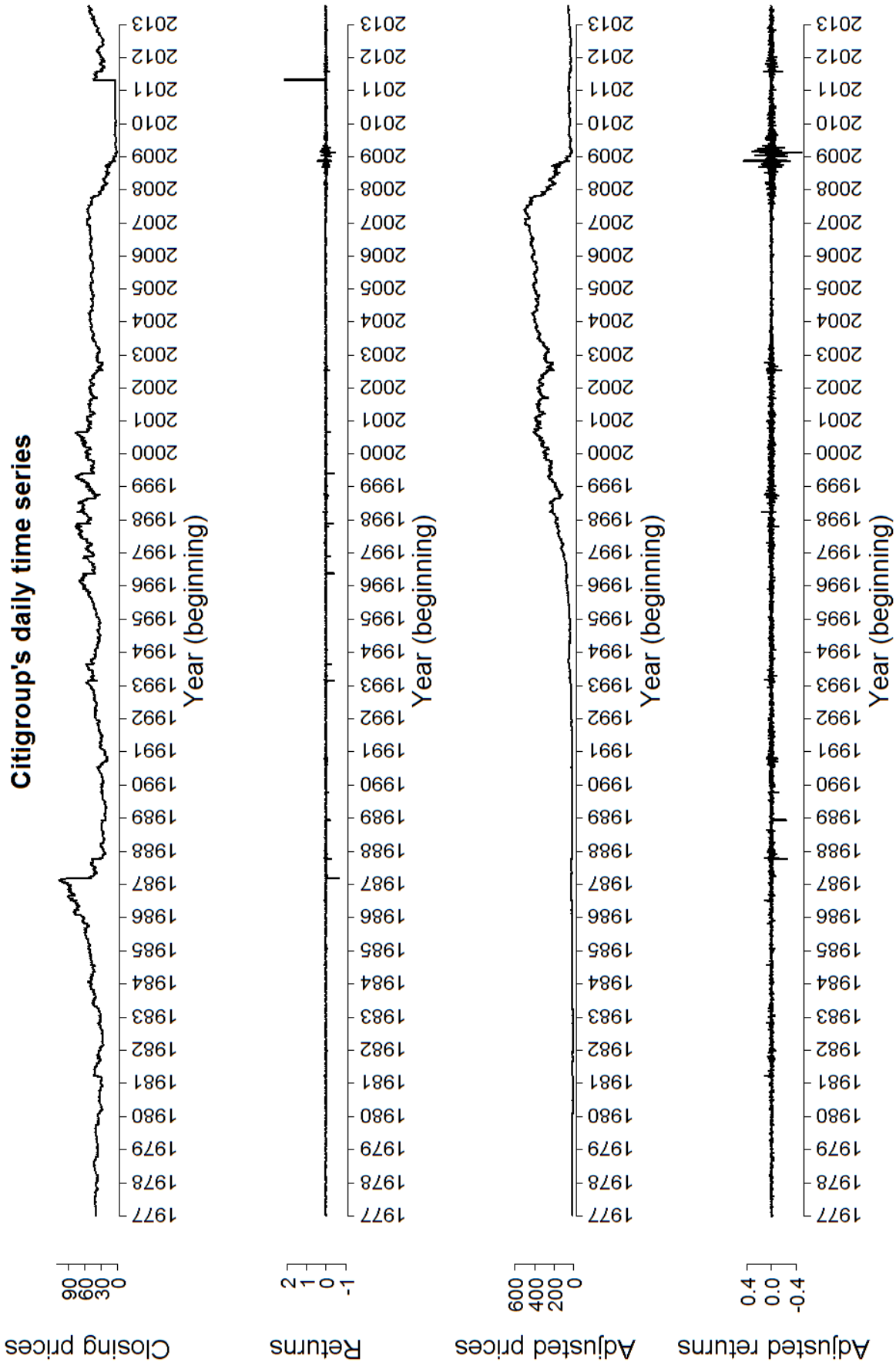


Figure 4.1 Time series plot of Citigroup daily data. From top to bottom: close price, returns (close-to-close), adjusted close price, returns (adjusted close-to-adjusted close). Data range from 03 Jan 1977 to 31 Jul 2013.

4.1.1.2.1 Asymmetric impact of returns on volatility

A stylized fact that is widely documented in quantitative finance researches is the negative correlation between stock returns and their volatility. This phenomenon is largely attributed to the changes in leverage ratio corresponding to changes in stock prices. In particular, a fall in stock price generally has a much larger impact on volatility than a rise of the same magnitude, since it causes the debt-to-equity ratio to rise, thus provoking equity investors' concern about their claim in the firm. We can illustrate this feature by plotting the stock prices against corresponding absolute returns (as a conventional proxy for volatility):

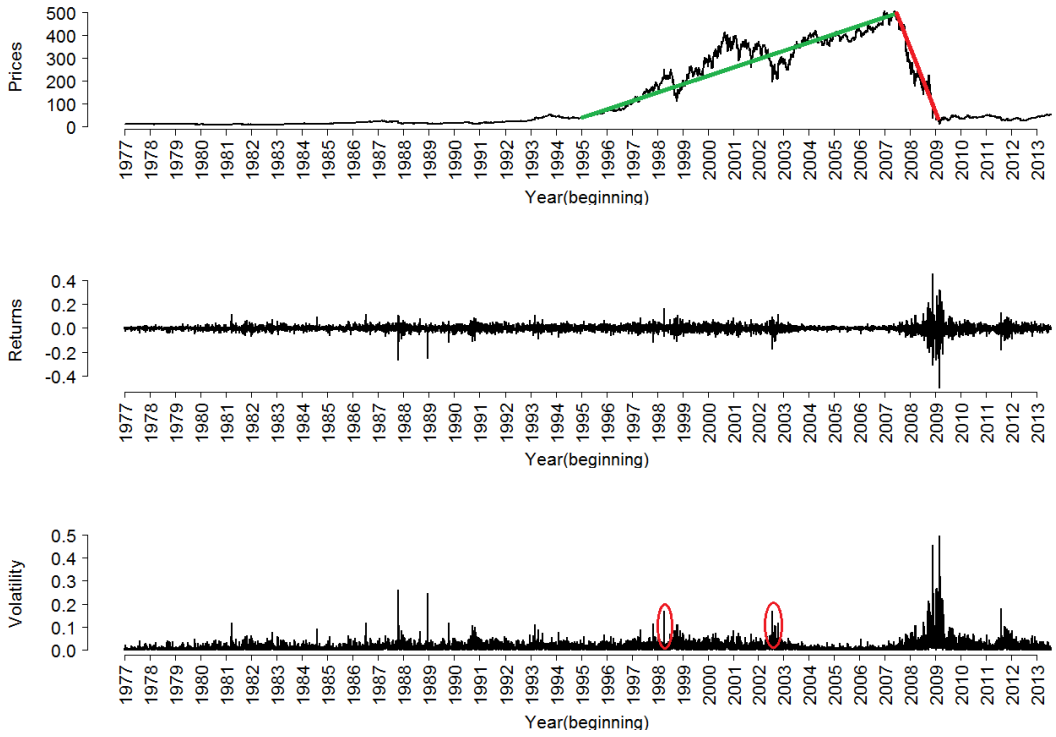


Figure 4.2 Comparison between Citigroup daily price series (adjusted for stock splits and dividends) and the corresponding absolute returns series, for the period 1977-2013. The two red ovals indicate periods when visible sharp price drops corresponding to high volatility (aside from the drop in the GFC).

For the period from 1977 to 2008 the increase in share price results in relatively low level of volatility that is distinct from that of the period from 2008 to 2010. From Figure 4.2 we can see clearly the inverse direction/sign exhibited by stock returns and volatility: whenever prices go down, volatility goes up and vice versa. Furthermore, the impacts of positive and negative price changes on volatility are dramatically different: a steep rise from 1995 to mid 2007 (indicated by the green trend line) results in a change of volatility that has roughly the same magnitude as the change corresponding to the period from mid 2007 to 2009 (in-

dicated by the red trend line). In other words, the strong fall during the three years of GFC undid the gain built up in previous 12 years and was undoubtedly related to one of the most volatile, unprecedented periods in the history of the finance world. Aside from this obvious feature, the asymmetric relation between price changes and volatility is repeated throughout Citigroup's timeline. This relationship was explicitly studied via the specification of asymmetric GARCH models earlier. It will be further elaborated at different time horizons in later sections via the investigation of the leverage effect hypothesis.

On the other hand, looking at Figure 4.2 one can argue that the relationship between the direction of price movement and the magnitude of volatility can be somewhat expected and the above argument about impact of the sign is not necessarily applicable. The reason is that if we consider volatility to be proportional to the absolute value of the quantity $\frac{\Delta P}{P}$ (where P is the price) then surely the volatility of period 2007-2009 is much greater than that of period 1995-2007. This is simply because this quantity is also proportional to the slope of the trend lines associated with these periods. The longer the period, the smaller the slope (in terms of absolute value) and vice versa.

This counter argument implies that what really affects volatility is not the direction of price movements, but their speed. If this point is valid, then the steeper (faster) the price goes up/down, the more volatile it will be, i.e. it must be true regardless of the direction of the price change. As a matter of fact, in general the bearish movements are associated with shorter time periods than that of bullish movements so that the bearish trend lines almost always have slope greater than that of bullish trend lines. In other words price decreases almost always at a faster rate than price increases. More evidence can be observed, for example, at the sharp price drops in late 1998 and mid 2002 corresponding to two periods of relatively high volatility (indicated by the red ovals in the last plot of Figure 4.2). On the other hand, we do not observe the reverse phenomenon, i.e. bullish trend lines having greater slope, thus without a 'counter-factor' we can not verify whether or not the direction of price changes is really unimportant. In any case, the observation of asymmetric impact of price changes remained relevant, at least in our sample.

4.1.1.2.2 Preliminary data analyses

Leptokurtic returns distribution 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 4.3.

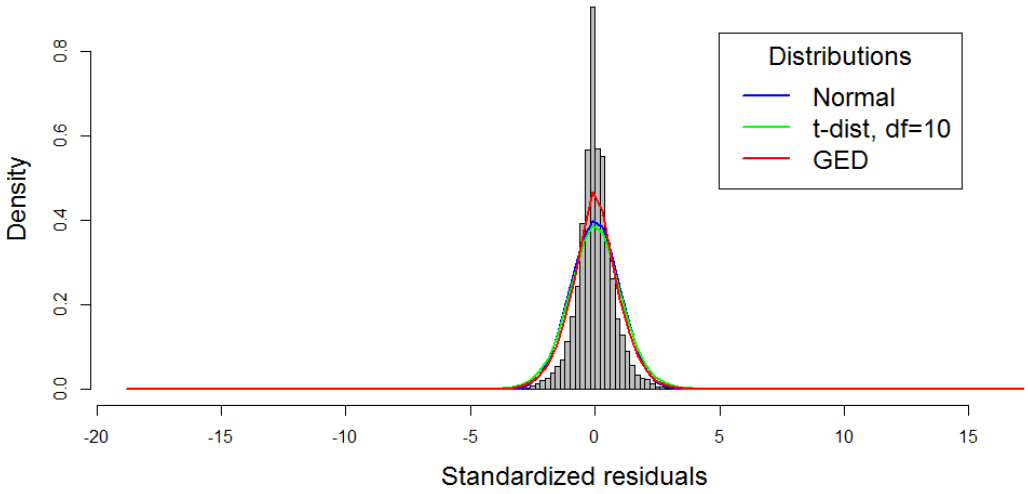


Figure 4.3 Histogram of standardized daily Citigroup returns for period 1977-2013, with lines indicating fitted normal, student and GED density functions superimposed.

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 4.4, when studying the boxplots of the 4 time series we can clearly tell the significant reduction of outliers when replacing the 'raw' data with winsorized data. 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.

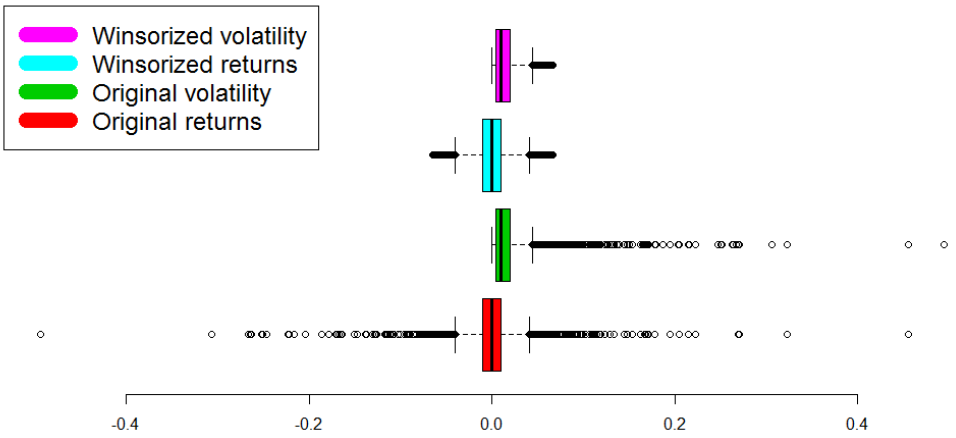


Figure 4.4 Boxplots of daily time series of Citigroup for period 1977-2013. The black line indicates the median. The number of outliers (or the length of the whiskers) is reduced significantly with winsorized data.
Here the width of the boxes represents the Interquartile Range (IQR), or the difference 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 et al. (1983)). Any value falling outside of the range implied by these whiskers is considered an outlier.

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 4.1. 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 normality assumption for all 4 time series, implying a heavy-tailed distribution. Therefore it seems reasonable to account for leptokurtosis with the specifications of our GARCH estimates (as shown in section 3.3.2, Chapter 3). 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). 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}	$\|r_{raw}\|$	r_{wins}	$\|r_{wins}\|$	r^2_{wins}
Mean	0.000189	0.0158056	0.0003052	0.014786	0.0004308
Median	0	0.0101181	0	0.010118	0.00010237
Variance	0.000703	0.0004537	0.0004308	0.000212	6.92×10^{-7}
Skewness	-0.605176	6.576009	0.0246598	1.639483	3.174645
Kurtosis	42.9454	84.3004	1.725833	2.58466	10.49374
JB	710028.5	2800229	1147.528	6706.388	57869.75
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
LB (21)	144.31	17018.05	52.3834	13161.6	14189.44
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table 4.1 Summary statistics for Citigroup daily data for the period 03 Jan 1977 to 31 Jul 2013. p-values are reported in parentheses. JB, LB(21) indicate the Jargue-Bera and Ljung-Box statistics, respectively.

Long-run autocorrelation When we examine Figure 4.1 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 4.5.

For the returns series, although it is difficult to observe the overall significance of the ACF, the Ljung-Box portmanteau test in subsection 4.1.1.2.2 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 as far as lag 400. This is consistent with our discussion of long-memory behaviour in Chapter 2. Furthermore, when we increase the lag range, we observe a more visible hyperbolic decaying pattern of the ACF functions of squared returns and absolute returns. This indicates a long-memory behaviour of these processes. This feature implies the long run impact of return shocks, which must be taken into account when modelling volatility.

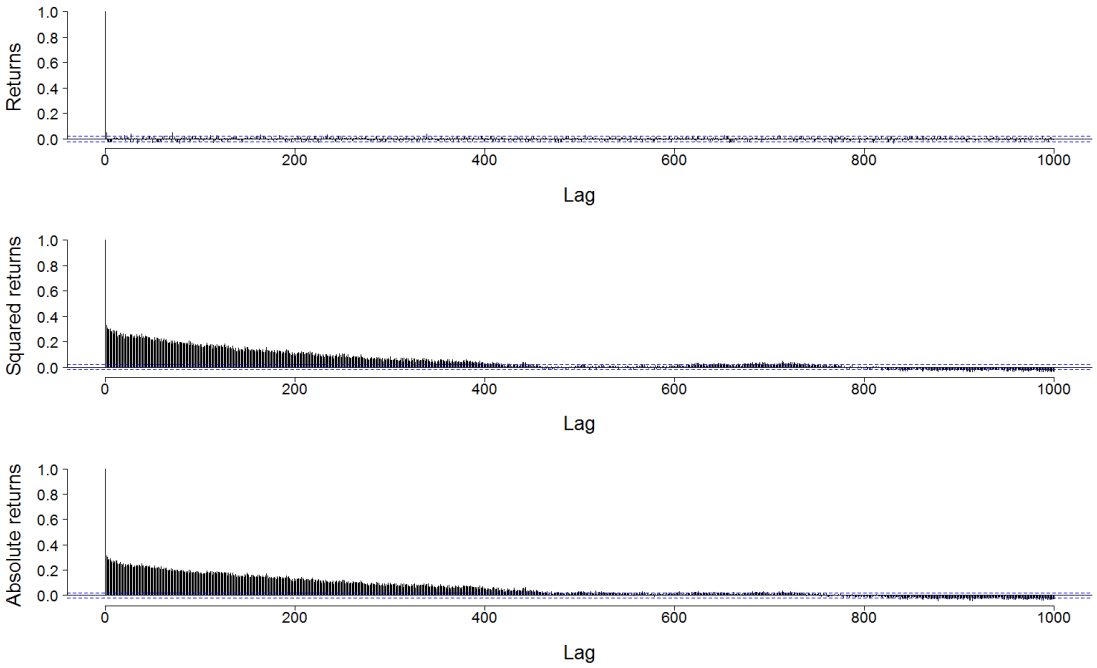


Figure 4.5 Correlograms of Citigroup daily time series for the period 1977-2013, up to 1000 lags. The blue dashed lines indicate the 95% confidence intervals.

Test for unit root non-stationarity It is observed that unit root non-stationary time series often exhibit slow decaying ACFs similar to those of stationary long-memory processes. Therefore it might not be possible to distinguish the two type of processes relying only on the ACF (Brooks (2002)). To find out whether our time series are non-stationary, we apply the Augmented Dickey-Fuller (ADF) test for unit root (see 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 up to order p serial correlation in the data generating process which could invalidate the statistical inference of DF test. The 'optimal' number of augmenting lags (p) is determined by minimizing the Akaike Information Criterion. ADF test statistic is the t-stat of the OLS estimate of γ. The null hypothesis of 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 4.2 the ADF test rejects the

null hypothesis at any level of significance for all series. We can conclude that returns, squared returns and absolute returns are stationary, thus validating our long-memory estimating methods discussed in Chapter 2 and Appendix B.

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

4.1.2 Intraday data

This section examines specific stylized facts of intraday returns and volatility, using a large sample of 5-minute returns of Citigroup throughout the year 2007. Its main goal is to identify representative characteristics of the underlying mechanism determining the relationship between returns and volatility on the one hand, and to clarify the dependence structure of volatility on the other.

In our analysis we obtain a tick-by-tick data sample from Sirca database (`http://www.sirca.org.au/`) for the full trading year of 2007, excluding weekends and holidays. We compute continuously compounded returns by taking logarithmic differences of the volume weighted price (denoted as VWP):

$$\mathrm{VWP}_f = \sum_{i=1}^{t} p_i \times \mathrm{Vol}_i$$

where f indicates the trading frequencies (i.e. 5 minutes), t is the number of ticks within that particular time frame. p_i and Vol_i are the trading price and volume at the i-th tick. As stated in Chapter 3 we omit the observations corresponding to the first and last ticks of the day when computing the VWP_f. This is because the volumes associated with these observations reflect adjustments to accumulative overnight information and closing session's activities, thus resulting in unusually high volatility. Thanks to the availability of tick-by-tick data, with this method we can significantly reduce the loss of information: we only have to exclude the first and last ticks, rather than the whole first 5-minute return of the day, as in many previous studies (see e.g. Bollerslev, Cai, and Song (2000) and Lee et al. (2011)) [25]. The returns series are then computed from the VWP by simple

[25] In addition, with tick-by-tick data, we can form an intraday time series of returns at any frequency of our choice, giving our study greater flexibility. Although we are mainly concerned

concatenation (similar to the method discussed in subsection 4.1.1.1). After these adjustments we are left with 19,500 observations of 5-minute returns, for a total of 250 days, with 78 observations per day (corresponding to the NYSE's standard of 6.5 hours trading period, from 9:30 E.S.T to 16:00 E.S.T [26]).

4.1.2.1 Issues related to microstructure effects

Non-synchronous trading This phenomenon refers to the well-known fact that in general the trades on exchanges such as NYSE are based on different trading frequencies for individual stocks. Conventional time series analysis examines fixed intervals, equally-spaced series, most notably the daily returns. As shown in section 4.1.1, daily returns are generally computed from a series of close-to-close prices. However, the last transaction price from which return is constructed may occur at varying time points from day to day. Therefore, naively imposing that we have a time series equally spaced with 24-hour intervals might mask the true returns' independence (Tsay (2001)) and lead to spurious estimates of long-range dependence structure of intraday data. As such, the 'effect' in this context is associated with the spurious serial correlation that is widely observed in time series of (high-frequency) financial returns.

A common symptom of non-synchronous trading is revealed by the fallacious lead-lag relationship between two independent stocks' returns. If one stock is traded more frequently than the other, it is highly likely that the effects of end-of-the-day macroeconomics news might be promptly reflected on the closing price of the former stock, but not on the latter's price until next trading day. We know that returns over a predetermined interval are computed based on the closing price or the price recorded at the last instance of each interval. Because of so-called "non-synchronicity" the closing instance of different securities varies greatly in a random manner. For the minority, frequently traded, highly-liquid stocks whose prices are adjusted almost continuously based on incoming news, observed returns is a good proxy for 'true' returns. However, for thinly traded stocks which make up the bulk of the market, returns based on price adjustments do not reflect updated information. By the time market makers re-adjust their

with 5-minute returns, examining higher frequencies such as 1-minute could provide us with a better understanding of volatility dynamics, as illustrated by Dacorogna et al. (2001). The major disadvantage we encounter is data limitation, as initially we could only retrieve one year of tick data from Sirca database. Longer sample of high frequency prices can be acquired from the Bloomberg data terminal hosted by the School of Economics and Finance of Victoria University of Wellington. Results obtained from studying these data could be presented in a future study.

[26] E.S.T: Eastern Standard Time, which is the time used in the Eastern Time Zone (North America) when observing Standard time (Winter and Autumn).

price (based on the signal of other stocks' prices) there is a delay. In the security microstructure literature, the difference between true returns and observed returns is more commonly referred to as the difference of a frictionless compared to a frictional world (See e.g. Cohen, Hawawini, Maier, Schwartz, and Whitcomb (1980)).

All in all, a misleading significant cross-correlation between the two returns series could be expected, even when the underlying data generating processes are independent, at least in such short intervals. This intuition may provide an explanation for the empirical lead-lag relationship between the returns of large and small firms (presumably the former are more frequently traded than the latter). Assuming the information quality given by large firms' price signal is greater than that of small firms, the covariance of present small firms' returns with past returns of large firms would be greater than the covariance of current large firms' returns and past small firms' returns (Chan (1993)). Analogously, a portfolio containing such assets is likely to induce significant lag-1 serial correlation, which is documented in the study of USD-DEM exchange rate by Dacorogna et al. (2001) and index returns by Chan, Chan, and Karolyi (1991).

When studying a single stock, an indicator signalling the presence of non-synchronous trading is the significant serial correlations of return series at lower lags, especially at lag 1 [27]. Cohen et al. (1980) pointed out several empirical phenomena regarding non-synchronous trading, one of which is weak serial correlation of single stock returns that has either positive sign (high value, heavily traded stock) or negative sign (thinly traded stock). Because we are interested in the dependence structure of volatility of a single security that is actively traded, it is necessary to be able to detect whether such problem exists in our data and address it accordingly. In other words, we focus on detecting and adjusting for this effect rather than modelling it.

Bid-ask bounce We know that for some stock exchanges, market makers are given the rights to decide on the prices at which they buy (bid) or sell (ask) stocks. The positive difference between these prices, called 'bid-ask spread', is a compensation granted to market makers for providing liquidity. According to

[27] For individual stocks, Tsay (2001) proved that there exists a significant negative first lag autocorrelation within the high-frequency return series : $\mathrm{Cov}(R_\tau, R_{\tau-1}) = -\pi\mu^2$ where $\mu = E(R_\tau)$, providing a non-zero probability (π) that the stock is not traded at a particular time period (i.e. when non-synchronous trading exists). This model is based on Lo and MacKinlay (1990)'s stochastic model to study this phenomenon, which was also known as the problem of 'non-trading' or 'stale-price adjustment' among financial economists. The latter term is a reference to the lagged adjustment of less frequently traded stocks. For a detailed review of several alternative models, see e.g. Campbell et al. (1997).

Tsay (2001), if the previous price takes the value of bid price, then the current price will take the value of ask price and vice versa. The problem is, this 'bouncing effect' could result in spurious volatility calculation at very high frequencies. Although it is widely claimed that increasing sampling frequency (of price changes, or returns) would help increase the accuracy of volatility estimates as standard deviation (See e.g. French, Schwert, and Stambaugh (1987)), this is not true at extremely high frequencies. For example, simply computing the standard deviation of prices moving up-and-down at, say 1 tick per second, would yield the same volatility as if the price has moved $\sqrt{3600} = 60$ ticks in an hour, while in fact there was no movement outside the range of the bid-ask spread. More importantly, similar to non-synchronous trading, this effect will induce a statistically significant first-lag serial correlation of the returns time series because of the spurious volatility at high frequencies. Similar observations are extensively explored by Campbell et al. (1997).

Remark Despite the issues related to microstructure effects, attributing the high frequency serial correlation solely to such effects was proved to be too restrictive and may lead to ignorance of 'genuine' autocorrelation that are the result of other price-adjustment factors (Anderson, Eomb, Hahn, and Park (2012)) (See also Atchison, Butler, and Simonds (1987), Lo and MacKinlay (1990) and Chan (1993)). In other words, these effects might only be a partial cause of market frictions. However, regardless of the cause, returns' autocorrelation poses a serious problem for traditional pricing models which are based on the assumption of some form of martingale price process, i.e. uncorrelated returns. Therefore it must be accounted for before undertaking further analysis.

Examination of our sample of 5-minute returns is consistent with the above arguments, showing the presence of a significant positive autocorrelation coefficient of 0.2143 at lag 1, as seen in Figure 4.6 [28]. We also note that the evidence of microstructure effects is more compelling when we use higher frequency returns, such as 1-minute returns. Following Andersen et al. (2000), we produce an adjusted return series by de-meaning and eliminating the first order serial correlation. As expected the adjusted returns series display much weaker serial correlation at lower lags and shall be used for all subsequent analyses in this

[28] Many researchers found 5-minute to be the "optimal" sampling rate with minimal impact of microstructure effects (See e.g. Campbell et al. (1997)). However, it appears that such an effect still presents in our sample. This observation is in line with the empirical findings of Cohen et al. (1980).

chapter, and will be denoted as r_{5min}:

$$r_{5min} = (r_\tau - \bar{r}) - \rho_1 \ (r_{\tau-1} - \bar{r}) \tag{4.1.1}$$

where $\rho_1 = 0.2143$, $\tau = 1, 2, ...19500$ and $\bar{r}$ is the sample mean return.

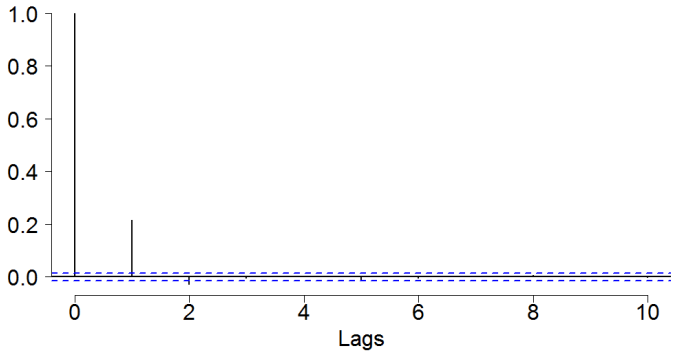


Figure 4.6 Correlogram of Citigroup's 5-minute returns (adjusted for microstructure effects). Sample size is one year, starting from 03 Jan 2007. The blue dashed lines indicate the 95% confidence interval.

4.1.2.2 Empirical stylized facts of 5-minute time series

Autocorrelation Preliminary analysis of the ACF of intraday returns adjusted for microstructure effects (r_{5min}) and its absolute value is illustrated in Figure 4.7. The returns series exhibits insignificant serial correlation. On the other hand, for volatility series, we can clearly observe the prominent periodicity in its autocorrelation sequences, with peaks corresponding to every one day (or at lags that are an integer multiple of 78).

There are additional important features about the volatility ACF that can be pointed out: First, it exhibits a very persistent behaviour: it stays well above the upper bound of the 95% confidence interval (calculated under a Gaussian null) until around lag 2,500, corresponding to a horizon of about one month. Second, similar to the ACF of daily volatility, when we ignore the daily seasonality the intraday ACF decreases hyperbolically, a stylized fact documented in numerous analyses (See e.g. Andersen et al. (2000), Gençay et al. (2002), and Lee et al. (2011)). Third, when taken together these two facts imply a long-memory volatility generating process, with a strong seasonal component.

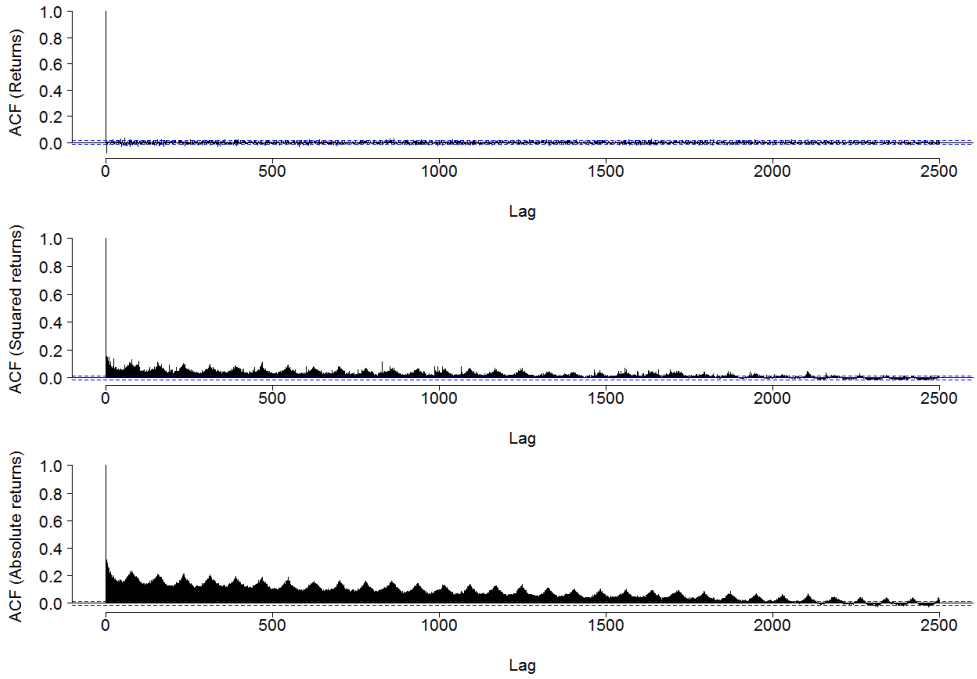


Figure 4.7 Correlogram of 5-minute returns, squared returns and absolute returns of Citigroup to up to 2500 lags corresponding to one month's time. Sample size is one year, starting from 03 January 2007. The blue dashed lines indicate the 95% confidence intervals.

Intuitively, the intraday seasonality is explained by the market trading activities throughout the day. According to Andersen and Bollerslev (1997b), among others, this type of systematic movements in volatility is largely attributable to profound interaction around market openings and closures, resulting in a particular 'U-shape' pattern of intraday volatility (thus explains why the volatility ACF exhibits repetitive 'U-shapes' occupying every 78 lags, as seen in Figure 4.7). To better clarify this point, following Andersen and Bollerslev (1997b), Bollerslev et al. (2000) and Lee et al. (2011), we construct the intraday volatility series averaging over the full sample of the year 2007 as:

$$\bar{\sigma}_i = \frac{\sum_{t=1}^{T} |r_{t,i}|}{T}$$

where T is the number of trading days in 2007 ($T = 250$) and $\bar{r}_i = \left(\sum_{t=1}^{T} r_{t,i} \right) / T$ is the averaged return at the i-th interval ($i = 1, \ldots, 78$). Consequently, the outputs can be thought of as series of 78 observations representing an 'average' trading day.

From the top left plot of Figure 4.8 we see that the values of average 5-minute return series for the year 2007 are mostly negative, indicating poor overall per-

formance in the eve of the global crisis. Interestingly enough, however, a certain degree of optimism from investors clearly results in the positive returns at the end of the day (say for the final trading hour or so). In line with Chang, Jain, and Locke (1995) and Andersen and Bollerslev (1997b), in the second left plot we observe the typical intraday 'U-shape' of the average volatility series: starting out at 0.2% in the morning then smoothly decline to around 0.1% at mid day and rise up to 0.13% at interval 75 (corresponding to 15:45 p.m). The subsequent drop in volatility is attributable to the closing of the cash market near the end of NYSE trading period (similar to an observation made by Andersen and Bollerslev (1997b)). We can see that the sequence of U-shape observed in the ACF (Figure 4.7) is merely an artifact of this intraday U-shape.

In addition, as we smooth the average volatility series using an MRA decomposition, the 'U-shape' becomes somewhat clearer at higher level. This is due to the smooth series being freed of high frequency fluctuations, retaining only the underlying trend of volatility movements. In particular, 'U-shape' is more visible from level 1 to 3, all of which show relatively lower volatility throughout the day compared to the beginning and ending sessions. Because we have skimmed the details information, the magnitude of the series decreases, thus when plotting on the same scale the 'U-shape' pattern dies out at level 6. In addition, the caveat that not much data have been used to plot these smooths (only 78 data points), also results in decreasing magnitude at higher levels.

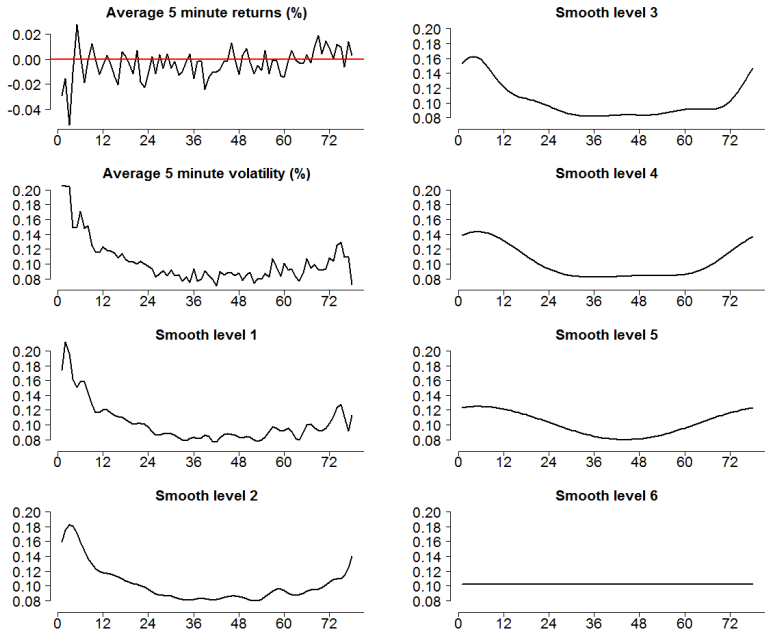


Figure 4.8 Average 5-minute returns ($\bar{r}_i$) and volatility ($\bar{\sigma}_i$) for the year 2007, together with smooth series of average volatility. All data are measured in percentage and plotted on the same scale. Every 12 five-minute intervals represents one hour of trading on the NYSE.

Seasonality and long memory Andersen and Bollerslev (1997b) assert that the dependency dynamics of intraday volatility could be distorted by the type of periodicity discussed above, thus resulting in fallacious implications when interpreting such dynamics without first accounting for the periodicity. As illustrated in section A.2.5 of Appendix A, the wavelet-based Multi-resolution Analysis (MRA) provides us with a tool capable of filtering out all intraday periodicities, by retaining the smooth series corresponding to a frequency just one level lower than daily frequency. Therefore in this subsection we perform a level 6 MRA on our 5-minute returns series (which has been adjusted for microstructure effects), because the level 6 smooth series is free of trading activities related to a frequency band ranging from $\frac{1}{2}$ (10 minutes period) to $\frac{1}{2^7}$ (640 minutes period), inclusive. It follows that this band contains all intraday activities, since we only have 78 intervals of five-minutes per day (corresponding to 390 minutes of trading) [29].

Consistent with the findings of Andersen et al. (2000), the filtered returns exhibit strong and persistent volatility dependency [30]. This is illustrated by the

[29] Note that if we stop at level 5, we will only be able to account for activities with a period of length up to 320 minutes (corresponding to a frequency of $\frac{1}{2^6}$) and thus leaving some intraday activities unfiltered.

[30] Similar to the long-memory interdaily pattern of the ACF observed from the bottom plot of Figure 4.5, subsection 4.1.1.2.2, as they represent essentially the same frequency.

ACF of the absolute filtered 5-minute returns in Figure 4.9 [31]. Furthermore, we can see that after filtering out the seasonality, wavelet transformation do a good job preserving the basic features of the ACF. This proves that the long-memory of daily volatility (now represented by the filtered data) is not just an artifact of high-frequency seasonality.

Henceforth we shall denote our seasonally adjusted intraday returns series as r_{5min}^{sadj}, as opposed to the original returns r_{5min} (which had been corrected for spurious serial autocorrelation). Summary statistics for these two series as well as for their absolute counterparts is reported in Table 4.3.

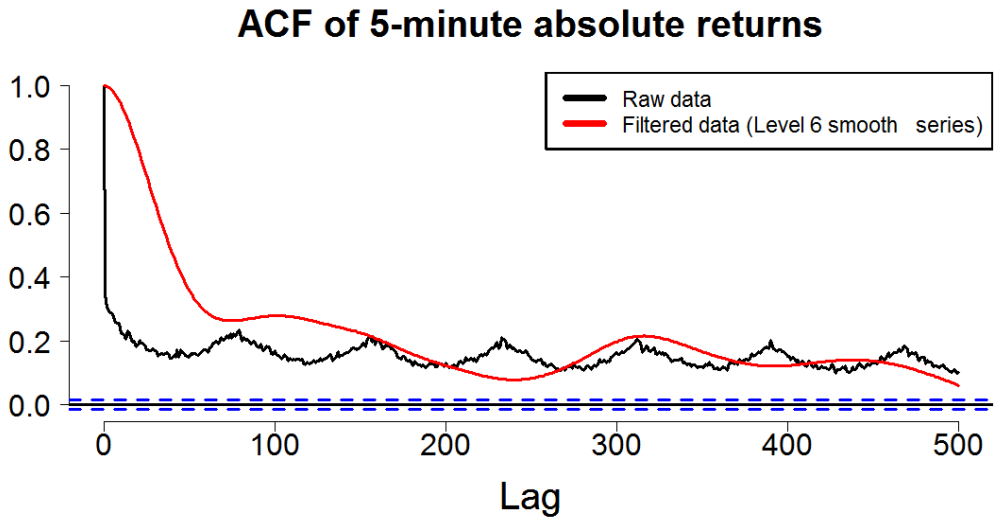


Figure 4.9 Correlogram of 5-minute absolute returns of Citigroup. Sample size is one year, starting from 03 January 2007. Raw data refers to the 5-minute data adjusted for microstructure effects. The red line indicates ACF of the absolute filtered returns (obtained from a level 6 MRA smooth series). The blue dashed lines indicate the 95% confidence intervals.

[31] For comparable experiments, see Gençay et al. (2002) (p. 148-149). These authors propose a filtered returns series formed via dividing the original 5-minute returns by the sum of wavelet details to get rid of all intraday fluctuations. However, we find using the smooth series is more intuitive and direct to obtain the same goal, provided that we are not using a multiplicative model.

	r_{5min}		$\lvert r_{5min}\rvert$		r^{sadj}_{5min}		$\lvert r^{sadj}_{5min}\rvert$	
Mean	-3.56×10^{-5}		0.001024		-3.56×10^{-5}		0.000148	
Median	-2.74×10^{-5}		0.000670		-2.44×10^{-5}		9.69×10^{-5}	
Variance	2.48×10^{-6}		1.43×10^{-6}		5.01×10^{-8}		2.94×10^{-8}	
Skewness	-0.15594		3.6659		-0.6369		3.340558	
Kurtosis	12.634		26.512		8.8582		16.410	
JB	129327	(0.0000)	612676	(0.0000)	64851.62	(0.0000)	254188.4	(0.0000)
LB (21)	925.2509	(0.0000)	23666.84	(0.0000)	350314.9	(0.0000)	334556.6	(0.0000)

Table 4.3 Summary statistics for Citigroup's intraday data for the period 03 Jan 2007 to 31 Dec 2007. p-values are reported in parentheses. (JB, LB) denote Jarque-Bera and Ljung-Box test statistics.

4.2 Main results

4.2.1 Introduction

In this section we shall demonstrate the use of Multi-resolution analysis discussed in Appendix A to investigate the relationship between returns and volatility. Throughout the following studies, MRA is extensively applied to both returns and volatility at the two 'original' frequencies described previously: daily and 5-minutes. The comparison of our studies at these frequencies will be showed to reinforce each other strongly. The universal approach will be based on a multi-scale/multi-frequency perspective, in hope of finding intuitive insights regarding volatility-returns dynamic.

The rest of the section shall be organized as follows: Subsection 4.2.2 serves as the starting point. It begins with the basic use of MRA on the whole daily data sample to give the overall impression of how returns and volatility interact at different scales. In particular, original data series will be decomposed into scale-specific details and smooths "components". As such, the details contain information associated with activities at high frequencies while the smooths act as "averaged" series containing information at the remaining lower frequencies. Subsection 4.2.3 explores how using these information may help increase the goodness-of-fit of the basic GARCH(1,1) model. Additionally, in this subsection we introduce the use of wavelet variance analysis, which is designed to capitalize on the variance preserving property of wavelet transforms. By studying the evolution of the signal's variance across different scales (for daily data as well as original and filtered intraday data), we are able to provide some intuition as to the results of fitting GARCH models. Moreover, the subsequent findings reveal important insights concerning the validity of two well-established theories that aim to explain the volatility-returns relationship, viz. the leverage effect hypothesis and the volatility feedback hypothesis. Subsection 4.2.4 revisits the examination

of long-run behaviour in our data. Again the multi-scale approach is adopted, with an emphasis on the varying degree of long-memory across scales.

Before investigating the MRAs, we need to re-emphasize the interpretation of wavelet decomposition levels and corresponding frequency bands. Specifically, following the examination of discrete wavelet filters discussed in subsection A.2.2.1, Appendix A, here we present a list of the first 8 level-specific detail series and associated trading periods (for both daily and intradaily data) in Table 4.4. For smooth series, it follows that at level j, whereas the detail corresponds to all activities associated with the frequency band $\left(\frac{1}{2^{j+1}}, \frac{1}{2^j}\right)$, the smooth corresponds to all activities associated with frequencies lower than $\frac{1}{2^{j+1}}$ (or with periods longer than 2^{j+1}).

Detail Level	**Scale length**	**Frequency band**	**Period band**				
			For daily data		**For 5-minutes data**		
			In days	In months	In minutes	In hours	In days
0 (original)	1	$0 - \frac{1}{2}$	1-∞		5-∞		
1	2	$\frac{1}{2^2} - \frac{1}{2}$	2-4		5-10		
2	4	$\frac{1}{2^3} - \frac{1}{2^2}$	4-8		10-20		
3	8	$\frac{1}{2^4} - \frac{1}{2^3}$	8-16		20-40		
4	16	$\frac{1}{2^5} - \frac{1}{2^4}$	16-32		40-80	0.67-1.33	
5	32	$\frac{1}{2^6} - \frac{1}{2^5}$	32-64	$\approx$ 1-2	80-160	1.33-2.67	
6	64	$\frac{1}{2^7} - \frac{1}{2^6}$	64-128	$\approx$ 2-4	160-320	2.67-5.33	
7	128	$\frac{1}{2^8} - \frac{1}{2^7}$	128-256	$\approx$ 4-8	320-640	5.33-10.67	0.82-1.64
8	256	$\frac{1}{2^9} - \frac{1}{2^8}$	256-512	$\approx$ 8-16	640-1280	10.67-21.33	1.64-3.28

Table 4.4 Frequency bands and period bands corresponding to first 8 level-specific wavelet detail series. For 5-minutes data, it follows that a full trading day of the NYSE only contains 390 minutes (or 6.5 hours).

4.2.2 Multi-resolution analysis

In Figure 4.10 we plot an 8-levels MRA for our daily returns and volatility time series. In Appendix A we show that the details and smooths of the MODWT are appropriately aligned with the original time series. The smooth series at level 8 reveals relatively stable returns and volatility (in terms of magnitude) before

the turmoil of the 2008 GFC. The most notable 'bulge' occurs at the height of the crisis, i.e. in 2008, and does not die out even at the highest level. We also note that since this is an additive decomposition, the sum of the details and the smooth equals the original series. To obtain a clearer impression we compare the MRA of returns and volatility on the same scale in Figure 4.11. Here we only plot the last three detail series and the smooth series, since the lower level series contain too many oscillations that would obscure the visual inspection. Note that details can be regarded as differenced series, while the smooth series is analogous to averaged series. Comparing the movements of both time series on the same scale reveals important features of their interaction.

The plots of details illustrate the relationship between scale-specific changes (difference) in returns and changes in volatility. This means a positive/negative value of the details indicate an relative increase/decrease in the original data. In terms of the positions of oscillations, relatively synchronous movements of the details of returns and volatility can be observed, that is, a change in returns seems to be accompanied by a lagged change in volatility. Additionally, the directions of fluctuations are reversed: a rise/fall in returns tends to correspond to a fall/rise in volatility. This further supports the negative relation between the two variables previously observed in Figure 4.2. It also constitutes evidence supporting the 'sign' aspect of a leverage effect.

As pointed out in Chapter 3, the 'size' aspect of leverage effect refers to the stronger impact of a decrease in returns to volatility, compared to an increase with the same magnitude. As shown in Figure 4.11, we can spot a generally asymmetric pattern: a fall in returns, when compared to a rise in returns, corresponds to a larger change in volatility (see, for example, the period 2007-2012 of the details level 8).

We can also see a pattern of lead-lag relationship between returns and volatility: if we compare the relative position of returns peak/trough to the immediate subsequent volatility trough/peak it appears that returns are 'leading' volatility, for a period of a few months. This implies that present change in returns has certain impact on future volatility. Nevertheless, this lead-lag relationship is fairly ambiguous, as at some points we can almost say that it is volatility that is leading, in which case the causal relationship is reversed. Furthermore, the alignment of the two series is vastly different at different levels. Also, aside from the overall lead-lag pattern there are a couple of 'anomalies' that can be pointed out:

- There is some tendency for returns peaks to occur almost simultaneously as volatility troughs: see e.g. periods 1997-1999 and 1987-1988 at level 7.

- There are instances where the two series move together, i.e. their movements have the same direction (rather than reversed): see e.g. periods 1981-1982 and mid 2009-mid 2010 at level 7.

Another observation we can make is the magnitude of the changes of volatility is mostly higher than that of returns, especially at lower levels. This is because in the long run returns series does not deviate significantly from zero and it quickly reverses to the mean level. Hence performing a differencing operation on it would create small coefficients. On the other hand, the larger detail coefficients observed for volatility are the result of persistent and abrupt changes of volatility 'clusters'. Again, here we can see the manifestation of long-memory in volatility, which is distinguishable from the short-run dependence structure of returns.

In any case, the visual impression provided by our MRA gives us some ideas about the extremely complex nature exhibited by the multiscale return-volatility dynamic, which happens to be one of the oldest and most hotly debated topics in the empirical finance literature. Later discussions shall provide some preliminary justifications via the interrelation of two well-known theories: leverage effect and volatility feedback effect.

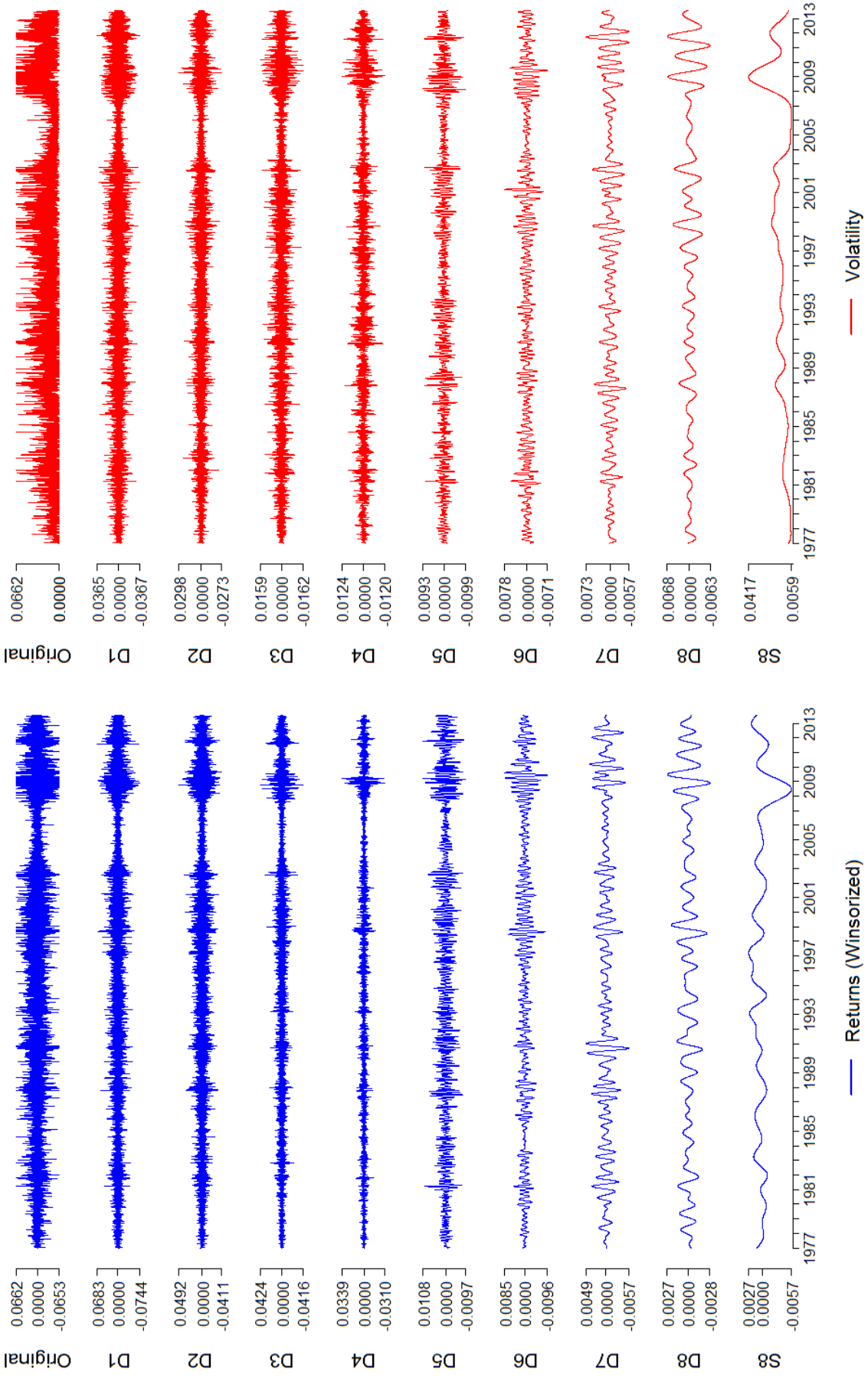


Figure 4.10 MRA of daily returns (left panel) and absolute returns (right panel). The decomposition is up to level 8. y-axis range indicates minimum and maximum values. Sample range is from 03 Jan 1977 to 31 Jul 2013.

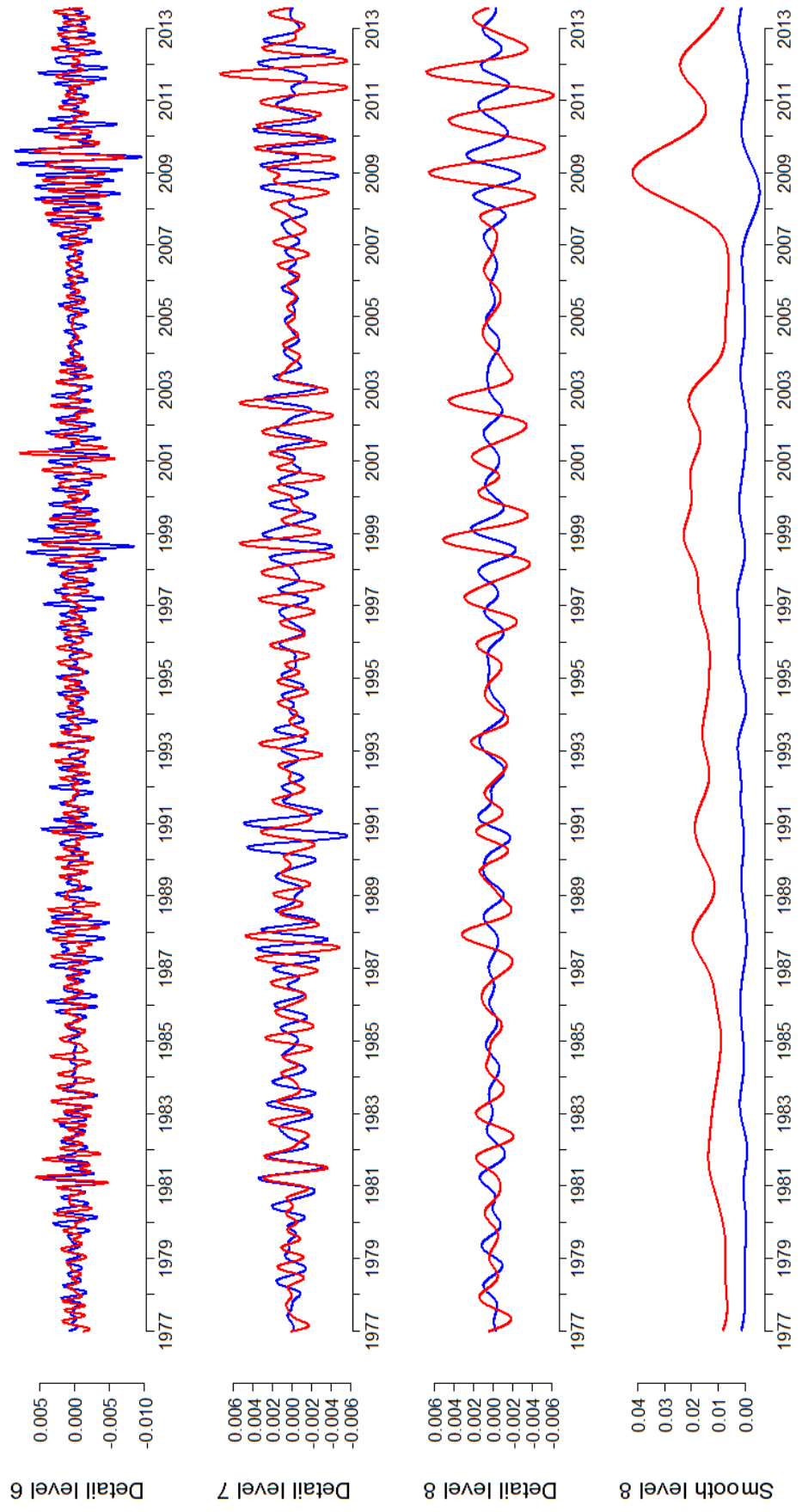


Figure 4.11 Comparing details level 6-8 and smooth level 8 of returns and volatility. Sample range is from 03 Jan 1977 to 31 Jul 2013.

4.2.3 Multi-scale GARCH estimates of daily conditional volatility

Previously in Chapter 3 we showed that GARCH(1,1) and related models seem to provide decent out-of-sample forecasts of conditional ex-post volatility by comparing these estimates with smoothed volatility series obtained from a wavelet-based decomposition. However, we were not be able to acquire a high degree of goodness-of-fit for the in-sample data. These observations give rise to the question: what happens if we change the input of these models, i.e. from returns series at original daily frequency to the MRA smooths (or its lower frequency representations)? The answer is that we may obtain different, and improved, degrees of goodness-of-fit. Therefore, this section aims to assess the ability of GARCH models in providing volatility estimates, by using level-specific smoothed returns series as input. To do this, the most direct approach is to be consistent with traditional GARCH specification, that is, we follow the framework discussed in Chapter 3. To simplify the experiment, we only present here the results for GARCH(1,1) with residuals assumed to have a Gaussian distribution.

4.2.3.1 Study design

Specifically, we first perform an MRA on the original (winsorized) daily returns at increasingly larger scales and extract the corresponding smooth series. Then we use these 'smoothed returns' as inputs to our GARCH(1,1) model and obtain level-specific $\hat{\sigma}_t$ estimates. Next, we compare the absolute values of smoothed returns (denoted as $S\left(r^j_{wins}\right)$ where j indicates the level) with the fitted $\hat{\sigma}_{j,t}$ at the same level. Finally, we specify the corresponding degree of goodness-of-fit with R-squared values obtained from a simple linear regression of $\left|S\left(r^j_{wins}\right)\right|$ against the corresponding $\hat{\sigma}_{j,t}$ specified by the following equation. Partly for ease of presentation and partly because it suffices for our analysis, we set our maximum decomposition level to 7:

$$\left|S\left(r^j_{wins}\right)\right| = a_j + b_j\ \hat{\sigma}_{j,t} + \epsilon_{j,t} \quad (j = 1, \ldots, 7) \tag{4.2.1}$$

In essence, the procedure is identical to our in-sample fitting as shown in Chapter 3, the difference being the multi-scale approach. That is, here the smoothed returns are in the position of the original returns. Results are shown in Figure 4.12. Note that for original data we obtain $R^2 = 0.2383$ instead of 0.1204 as reported in Section 3.3.1, Chapter 3 because at the moment we are using the 'winsorized' returns (also note how the absolute returns series looks 'trimmed' in

the top left plot of Figure 4.12 !). Apparently, excluding outliers in our sample gives us a better fit at the original level. With all that said, what we obtained from higher levels fitting is even more remarkable than this initial improvement.

The comparison across scales will give us an idea of how well our GARCH model performs at lower frequencies. To our knowledge, researchers with the same goal typically compare results at a few different frequencies (i.e. weekly vs. daily vs. monthly). Our approach is advantageous in that the comparison is more thorough. However, one setback is that the wavelet-based MRA can only account for fixed bands of frequencies rather than specifying individual frequencies. In addition, the leakage introduced by finite filters distorts the ideal frequency band and imposes a certain degree of imprecision (see section A.2.2 of Appendix A). Nevertheless, we can draw more comprehensive conclusions concerning the systematic choices of investors that vary at frequencies that were not documented in previous studies. In any case, to practitioners, deriving implications for bands of frequencies may be more reasonable and perhaps more realistic, since there is some 'fuzziness' about the perception of frequencies in practice (for example, the number of days in a month, or business weeks in a year, or even the number of business days in a week during holiday periods are all highly variable). Ultimately, the design and/or interpretation of empirical studies depend entirely on the goals of the researchers.

4.2.3.2 Results

The most striking observation is the immediate and significant increase of R^2 when we change from inputting the original data into GARCH to inputting the smooth data at higher levels. More importantly, R^2 is maximized at level 4 (which does not yield a significantly higher value than level 5) [32]. Following arguments similar to those of Section 3.4.2, Chapter 3, this evidence supports the importance of trading activities at frequencies lower than $\frac{1}{64}$ (corresponding to a period longer than 2 months) [33]. Furthermore, exclusion of the detail information corresponding to a trading horizon between 2 and 4 months results in a decrease of R^2 (from level 5 to 6). To be more clear, if we denote the level 6 detail returns as $D\left(r^6_{wins}\right)$ then we can express the relationship between level 5 and 6 smoothed

[32] We also find that with the increase of wavelet scale length, the value of maximized likelihood function remains unchanged, but do not pursue this issue further.

[33] Refer to the interpretation of Table 4.4.

returns as follows:

$$\begin{aligned} S\left(r_{wins}^{5}\right) &= \sum_{j=6}^{\infty} D\left(r_{wins}^{j}\right) \\ S\left(r_{wins}^{6}\right) &= \sum_{j=7}^{\infty} D\left(r_{wins}^{j}\right) \\ S\left(r_{wins}^{5}\right) &= D\left(r_{wins}^{6}\right) + S\left(r_{wins}^{6}\right) \end{aligned} \tag{4.2.2}$$

To sum up, our argument is that $S\left(r_{wins}^{5}\right)$ is most closely fitted by GARCH(1,1), and what makes it so important is $D\left(r_{wins}^{6}\right)$.

Conclusion Previous studies (e.g. Andersen and Bollerslev (1998)) imply that GARCH models provide decent estimates of the latent volatility process (proxied by a realized measure). Our results indicate that GARCH(1,1) model performs best with the inclusion of activities related to period lengths falling between 2 months and 4 months, in which variation of GARCH estimates was able to explain an astounding 93.27% of variation in empirical (smoothed) volatility. This is equivalent to saying that, if we use the GARCH estimates as an indicator, the level 5 smooth series (including all periods longer than 2 months) appears to be our best approximation to the 'core' component of the true data generating process of volatility. Of course we are not using our realized measure of volatility in this analysis, instead just opting for absolute returns as a benchmark to assess volatility estimates. Therefore, we also need to be careful when interpreting these results.

In any case, it is the low-frequency activities, not day-to-day activities (which are excluded via the filter of high-frequency information contained in lower levels), that account for the majority of the variation in volatility. This is consistent with the notion that in the long run, big price movements (which are associated with low frequencies and long-term traders) contribute the most to the volatility process and have a big impact on market. Analogously, the dominant influence on returns and subsequently volatility most likely arises from large, long-term economic cycles.

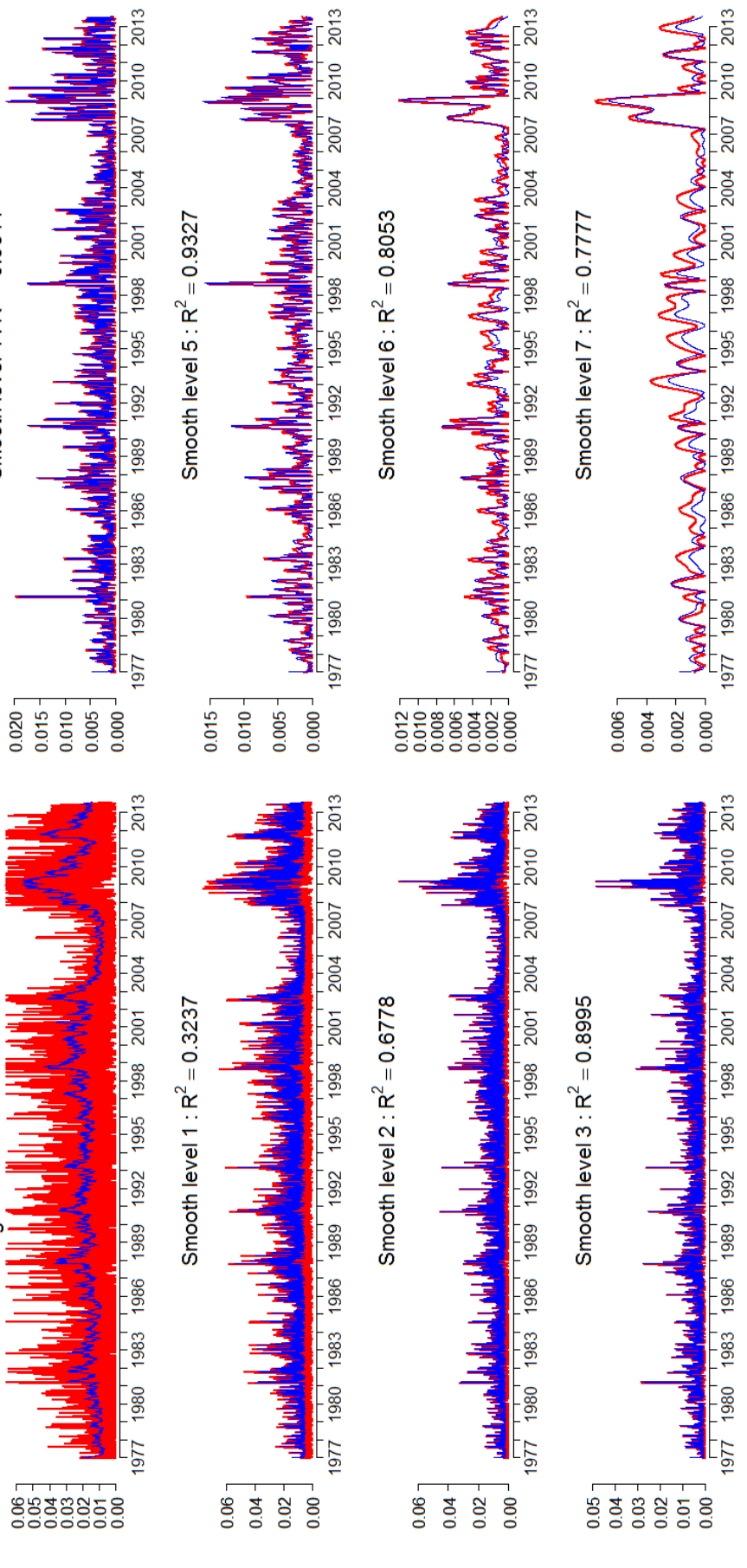


Figure 4.12 Comparing GARCH(1,1) fitted conditional volatility (red lines) with absolute smoothed returns (blue lines) at multiple scales. Here the smoothed returns obtained from MRA at each level are used as input for GARCH model.
The decomposition is up to level 7. Sample range is from 03 Jan 1977 to 31 Jul 2013. R-squared obtained from regressing $\left|S\left(r^j_{wins}\right)\right|$ against $\hat{\sigma}_t$ is reported.

4.2.3.3 How persistent is the leverage effect?

As pointed out at the end of Chapter 3, one direct evidence of leverage effects observed in daily data was the negative sign of the cross-correlation of returns and volatility at positive lags [34]. We also show that the number of significant coefficients is low when considering the in-sample data (i.e. prior to the 2008 GFC) but it is higher when we use the out-of-sample data, which supports a stronger effect at the beginning of the GFC. In addition, via asymmetric GARCH specifications we are able to quantify this effect for the pre-crisis period. In the following study we use a more complete sample range including all daily data available from January 1977 to July 2013.

4.2.3.3.1 Visual inspections

Interestingly, the CCF at original frequency reveals a stronger, more profound leverage effect (compared to the in-sample results), as can be seen in the top left plot in Figure 4.13. To interpret this CCF, we emphasize that the 'base' or the 'lead' series is returns, which means coefficients at positive lags imply an impact of past returns shocks on future volatility, while coefficients at negative lags imply effect of past volatility on future returns. Hence, a significantly negative coefficient at a positive lag is an indicator of the presence of leverage effect at that particular lag.

As we compute the CCFs of smoothed returns and smoothed volatility obtained via MRAs at longer scales (up to 5 levels), the effect becomes stronger as well as more significant. Analogously, the long-run impact of present (negative) returns shocks to future volatility is expected to be much stronger when being viewed from a long-run perspective. For example, let us compare the relationship between a present shock in return and a future change in volatility 20 periods (lags) later, at two specific levels: 2 and 5. Recall that a smooth series at level 2 is associated with a trading frequency lower than $\frac{1}{4}$, in other words we are only concerned about transactions that happen every n days where $n > 4$. Now, from the second left plot in Figure 4.13 we can say that a 1% decrease in current returns would result in an increase of 0.112% in volatility 20 days later (since the

[34] According to Engle and Ng (1993), there is an ongoing debate as to whether change in leverage really has an asymmetric impact on the properties of returns' variances. On the other hand, in some cases changes in returns and volatility might have no connection to the actual changes in a firm's capital structure. Therefore, the term 'leverage effect' has been used largely as a conventional practice at best. For our purposes, we only focus on the 'sign' (i.e. negative correlation) instead of the 'size' (i.e. asymmetric impact) aspect of the phenomenon. As such, in this paper the 'leverage effect' may be simply understood as the negative returns-volatility relationship. In addition, to simplify the study we assume that this relationship is solely attributable to this particular sense of 'leverage effect', although this may not be the case (See e.g. Duffee (1995) for arguments regarding other factors influencing this relationship).

sample cross-correlation at lag 20 is $\hat{\rho}_{20} = -0.112$). However, a shock with the same magnitude would increase future volatility to an extra 0.288% when considering the level 5 smooth. This is consistent with the intuition supported by Lee et al. (2011), that at higher levels, or longer time horizons, investors have more time to fully interpret the risk associated with a high leverage ratio. Intuitively, as time passes by, more information regarding the company's performance could be gathered and incorporated into its stock price, thus consolidating market reaction.

In addition, we also observe more and more significant coefficients at negative lags as the level increases. This is because as stock returns become volatile due to previous shocks in returns, it is more likely that the volatility, as a measure of the firm's financial vulnerability, will build up momentum and cause further price drops. As discussed later, this 'reversed' causal relationship might be supportive of an alternative theory known as the "volatility feedback effect". In any case, it seems the presence of the leverage effect, as shown in Figure 4.13 is clearly dominant, in that the negativity of CCF coefficients at positive lags is much greater than that at negative lags.

To sum up, our data exhibit evidence that better market awareness at lower frequencies tends to strengthen the negative relationship between movement of returns and volatility. All things considered, perhaps it is not unreasonable to attribute this asymmetric pattern permeating multiple levels to a stylized fact of returns-volatility dynamic.

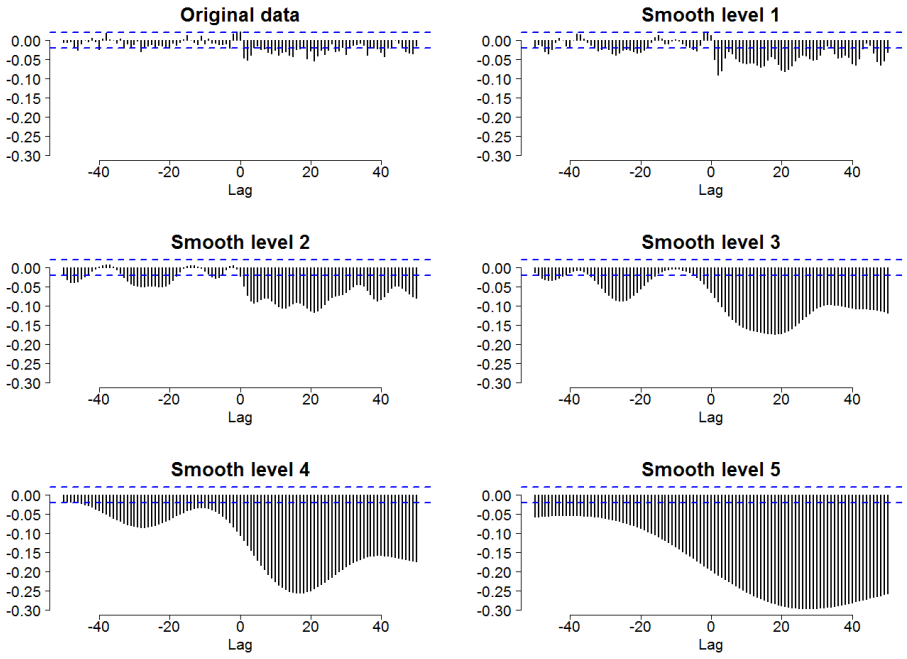


Figure 4.13 Multi-scale cross correlograms of the smoothed daily returns and smoothed volatility. The decomposition is up to level 5. Sample range is from 03 Jan 1977 to 31 Jul 2013. Blue dashed lines indicate the 95% confidence interval. All data are plotted on the same scale. Negativity at positive lags supports the negative impact of present returns shocks on future volatility.

Original frequency of leverage effect With modern advancement in information sources as well as information gathering mechanisms, it would be reasonable to assume that the market as a whole could react to a firm's slightest 'commotions' within a time frame less than one day, that is, the leverage effect should be perceived at an intraday level. This means we assume the validity of the EMH even in very high frequency trading.

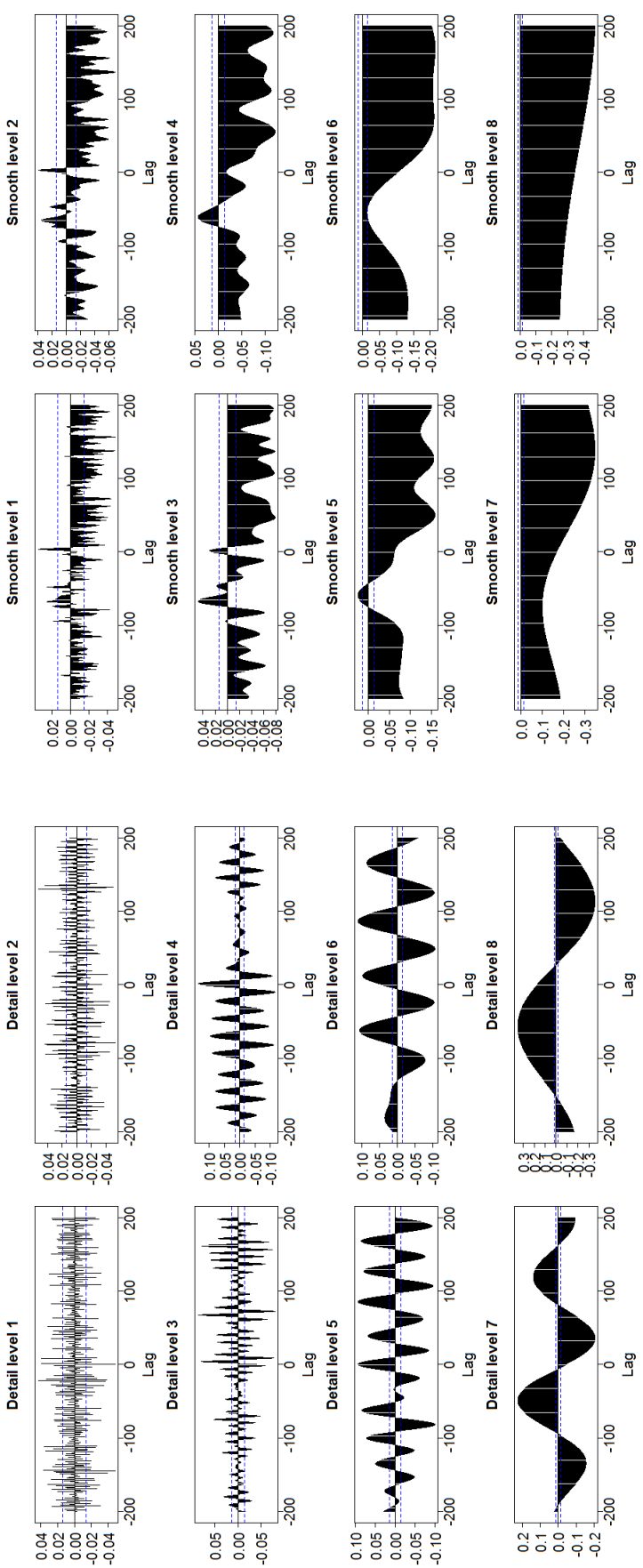
When investigating the original frequency at which the leverage effect is perceived, Lee et al. (2011) show that this effect is also observed in intraday data, but only at horizons longer than 20 minutes (i.e. at levels higher than 2). Their evidence implies that changes of returns within short periods of time such as every 5, 10 or 20 minutes will not have a significant impact on the perception of the risk associated with a raise in leverage ratio, even when such risk already exists. In particular, they used a test first proposed by Duffee (1995), which is tailored to be consistent with our notations as follows:

$$|r_{5min,t}|^j - |r_{5min,t-1}|^j = \alpha^j + \beta^j r^j_{5min,t} + \epsilon^j_t \tag{4.2.3}$$

where $r^j_{5min,t}$ and $|r_{5min,t}|^j$ are the scale λ_j components of the original returns and volatility, respectively. Given a significant negative estimate of β^j, the change in volatility is negatively related to change in returns, thus supporting leverage effect

at scale λ_j. It is worthwhile noting that these authors compare detail coefficients of returns and volatility at each level and only study the smooth series at the highest level. That is, given $j = 1, \ldots, J$ where J is the highest level of decomposition, we have the highest level smooth series when $j = J$. This is the standard approach based on MODWT and MRA, which is very straightforward for the purpose of studying the individual contribution of high-frequency components to the evolution of the signal as a whole, as discussed. Undoubtedly this approach provides us with valuable information regarding what makes the signal 'unique' at each level. However, we consider this approach inappropriate because the details are not representative of the underlying trend of original series and only reflect higher-frequency 'differences', or fluctuations. Therefore, it is not surprising to find little evidence of leverage effect when comparing detail returns and detail volatility. To investigate the linear relationship between two variables, as in this case, we believe it is crucial to study the relationship between their trend-following smooth series on a comparable basis. To do this, we first perform an MRA at each level, then plot the CCF of the corresponding smooth returns and smooth volatility at that level. For comparative purposes, we also repeat the experiment using the details, similar to what Lee et al. (2011) did.

As it turns out, analyses of detail coefficients show a relatively 'symmetric' pattern, where the cross correlation function takes on positive and negative values periodically, at both positive and negative lags as shown in Panel A of Figure 4.14. Thus it is very difficult to spot visual evidence of leverage effect even at the low frequencies (which makes Lee et al.'s findings questionable). In sharp contrast, the multi-scale correlograms between smoothed returns and smoothed volatility exhibit significant leverage effect even at the highest frequencies. Also it is easy to note the closely similar asymmetric patterns of intraday data in panel B, Figure 4.14 and daily data in Figure 4.13.



(a) Panel A: detail series

(b) Panel B: smooth series

Figure 4.14 Multi-scale cross correlograms between the 5-minute returns and volatility. The decomposition is up to level 8. Sample range is from 03 Jan 2007 to 31 Dec 2007. Blue dashed lines indicate 95% confident intervals.

4.2.3.3.2 Quantitative tests

Intraday data To verify the visual implications of Figures 4.13 and 4.14 discussed above, we first replicate the test specified by Equation 4.2.3 using both details and smooths of the original 5-minute returns series (r_{5min}) as regression variables. Results are reported in Table 4.5. Here we assign the subscript j to regressions using detail series and J for regressions using smooth series. We also compute the ratio between wavelet variance at each level and the total variance of returns based on the wavelet variances formulated in section A.2.3, Appendix A [35]. As we can see, the coefficients of the OLS estimator of β^j and β^J are significant at most of the levels (with the exception of the level 1 and level 5 smooth series). Taking statistical significance into account, we can say that for detail series, leverage effect is supported at levels 1,2,7,8 while it is only significant at level 6,7,8 for smooth series (although there is insignificant negative β^J for smooth level 1 and 5). The original 5-minute returns, as a combination of 9 scale-based components (8 details and 1 smooth components), also exhibits significant evidence of leverage effect.

In contrast to test results documented by Lee et al. (2011), our estimates for detail series imply that leverage effect only exists at the lowest and highest levels, whilst disappearing at intermediate levels such as 3,4,5,6. Results for smooths are somewhat in agreement with their findings, although the leverage effect is missing at all the first 5 levels, instead of only level 1 and 2. In addition, the joint significance of the coefficients in our smooth regressions is questionable as the R^2 values are much lower than that of the detail regressions. While the small R^2 obtained from detail regressions are comparable to that of Lee et al. (2011) (p.31, Table 6), results for smooth regression suggest no linear relationship between the regressors. Very small coefficient of determination (i.e. R^2 much less than 1%) is almost certainly attributable to the extraordinarily large sample size of 5-minute returns. As a plausible explanation, it is very difficult to find significant correlation of extremely long numerical vectors.

The main differences between our findings and those of Lee et al. (2011) include:

[35] Lee et al. (2011) pointed out that the OLS estimator of β for the original series can be expressed as a weighted sum of the estimators of β^j at different levels, in which the weight is the ratio of the wavelet variance at scale λ_j to the total variance. In other words: $\beta = \sum_{j=1}^{J} \left[\frac{\tilde{\sigma}^2(\lambda_j)}{\mathrm{Var}(r_t)}\right] \times \beta^j$, so that the returns-volatility relationship at original frequency can be interpreted as the weighted sum of same relationship at different levels.

- With respect to the results for details, it is not clear as to why the volatility-returns dynamic is in agreement with leverage effect hypothesis even at low levels for Citigroup's stock, whilst this is not true for the market index. Furthermore, why does the leverage effect hypothesis holds at intermediate levels for the index series and not for Citigroup's returns? Could the cause of these differences lie in the preference for various trading horizons of investors? Possibly, as traders would likely pay more attention to a stock's slightest movements (especially when they are speculating on it), while keeping in mind that big moves in a market index only occur over longer periods of time. The answer could also lie in the significant role the firm played during the turbulent times of 2008-2009, as the major catalyst facilitating the chain effect of the subprime mortgage loans fiasco.

- Now consider the results for smooths: assuming the reflection of risk associated with higher leverage ratio is postponed at high frequencies, then exactly at which frequency would the market start incorporating this information into stock prices? Will it be at every 20-minutes or longer as suggested by Lee et al. (2011), or at every 320-minutes or longer as implied by quantitative test results?
  In summary, panel B of Table 4.5 suggests that the leverage effect only exists for smooth level 6 or higher (corresponding to a trading period longer than 320 minutes). We already know that for our sample a trading day consists of 390 minutes, therefore it can be said that we find virtually no significant intraday leverage effect with the regressions of smooth data. This is altogether conflicting with Lee et al. (2011), the regression results for detail data and the plots in panel B, Figure 4.14.
  Granted, we only examine a year of 5-minute returns of a single stock, whereas the data sample studied in Lee et al. (2011) covers an 8 years period of the S&P500 index. The huge data discrepancy as well as the different nature of the signal compared to the previous study therefore are important to understand our differences.

Daily data Earlier we note that studies at level 6 for five-minute data is equivalent to studies at level 1 for daily data (as level 6 marks the transition from intraday to daily activities). Given the fact that for 5-minute data, at level 6 we observed leverage effect when using smooth series whereas using detail series yield no such evidence, it is necessary to repeat the test with daily data. If our intuition about the delayed incorporation of leverage risk perception is reasonable, we would expect to see a negative β at the daily level and lower frequencies. To

proceed we replace the 5-minute returns and volatility in Equation 4.2.3 with daily series (i.e. $r_{wins}^{j,t}$ and $|r_{wins}^{j,t}|$). Corresponding results are reported in Table 4.6. The table reveals several differences compared to its intraday counterpart:

- We observe strongly significant estimates of both β^j and β^J at all levels.
- Evidence of leverage effect exists at all levels except for the first level of detail series. Though the results for the first level details is not expected, this reconfirms the link between level 6 of 5-minute data and level 1 of daily data: for 5-minute data, level 6 detail coefficient is positive while level 6 smooth coefficient is negative. Since negative daily returns-volatility relationship is extensively documented in a vast number of previous studies, this result illustrates the fault of using detail series.
- In line with the top left plot of figure 4.13, the original daily data do not yield a conclusive evidence of leverage effect, with an insignificant positive β.

Remark In conjunction with Lee et al. (2011), we find evidence of leverage effect at all scales for daily data. However, for 5-minute data, this effect is missing at the first 5 levels instead of only the first 2 levels as these authors proposed. Specifically, a positive relationship is expected to exist with periods between 5 minutes and 320 minutes (corresponding to level 1-5 of the 5-minute smooths) and a negative relationship exists with longer periods. This basically means that there is no significant intraday leverage effect. In any case, leverage effect is strongly supported at any frequencies lower than daily.
The bottom line is that given the complex nature of return-volatility dynamic, as exhibited by our conflicting evidence, investors should adopt a cautious perception of the interaction between returns and volatility and the possible leverage effect, with respect to their trading horizons.

Panel A: Details								Panel B: Smooths				
Level	α^j ($\times 10^{-8}$)		β^j		R^2	$\tilde{\sigma}^2(\lambda_j)(\times 10^{-9})$	$\frac{\tilde{\sigma}^2(\lambda_j)}{\text{Var}(r_t)}$	Level	α^J ($\times 10^{-8}$)		$\beta^J(\times 10^{-3})$		$R^2(\times 10^{-4})$
D1	-5.0577	(0.9950)	-0.0449****	(0.0000)	0.001554	1158.548	0.4904						
D2	1.6197	(0.9958)	-0.0115*	(0.01846)	0.000285	647.330	0.2739	S1	16.2340	(0.9652)	-0.045731	(0.99005)	8.005×10^{-5}
D3	2.8932	(0.9806)	0.0116****	(0.0000)	0.000884	291.645	0.1234	S2	14.5972	(0.9197)	10.9082****	(0.0000)	13.865
D4	3.2600	(0.9479)	0.0149****	(0.0000)	0.003954	136.999	0.0579	S3	11.7083	(0.8533)	8.1433****	(0.0000)	19.415
D5	2.3265	(0.9170)	0.0022*	(0.04963)	0.000198	61.711	0.0261	S4	8.4580	(0.7774)	1.85612*	(0.03904)	2.1921
D6	1.0653	(0.9338)	0.0129****	(0.0000)	0.009774	30.642	0.0129	S5	6.1360	(0.6890)	-1.06475	(0.09253)	1.4565
D7	0.4061	(0.9031)	-0.0087****	(0.0000)	0.042545	19.952	0.008443	S6	5.0815	(0.4365)	-8.05245****	(0.0000)	2.3625
D8	0.2981	(0.8525)	-0.0097****	(0.0000)	0.097742	8.3745	0.003543	S7	4.6742	(0.3562)	-7.31255****	(0.0000)	1.4043
S8	4.3727	(0.3263)	-0.005175****	(0.0000)	0.003920	7.1690	0.003414						
Original series	1.1207×10^{-7}	(0.9906)	-2.0170×10^{-2} **	(0.00121)	0.0005383	2363.096	1						

Table 4.5 Results of the test for leverage effect in 5-minute data. Model tested: $\sigma_t^j - \sigma_{t-1}^j = \alpha^j + \beta^j r_t^j + \epsilon_t^j$.

$\tilde{\sigma}^2(\lambda_j)$ denotes the wavelet variance at scale λ_j. $\frac{\tilde{\sigma}^2(\lambda_j)}{\text{Var}(r_t)}$ denotes the ratio of wavelet variance at scale λ_j to the total variance of returns.

p-values are reported in parentheses.
Statistical significance indicators: * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$).
"($\times 10^{-i}$)" indicates measurement unit that should be multipled by the reported number to get the actual number.

Panel A: Details								Panel B: Smooths				
Level	α^j ($\times 10^{-8}$)		β^j		R^2	$\tilde{\sigma}^2(\lambda_j)(\times 10^{-6})$	$\frac{\tilde{\sigma}^2(\lambda_j)}{\text{Var}(r_t)}$	Level	α^J ($\times 10^{-6}$)		β^J		R^2
D1	35.9959	(0.9980)	0.04652****	(0.0000)	0.001791	202.366	0.468						
D2	-8.5086	(0.9988)	-0.02587***	(0.000158)	0.001545	113.136	0.262	S1	8.05445	(0.9055)	-0.0251****	(0.0000)	0.00297
D3	-11.6733	(0.9957)	-0.02830****	(0.0000)	0.007186	63.352	0.146	S2	6.42872	(0.8043)	-0.0195****	(0.0000)	0.00607
D4	-7.3270	(0.9928)	-0.00498*	(0.010747)	0.000705	27.901	0.065	S3	5.27880	(0.5915)	-0.0153****	(0.0000)	0.01173
D5	-15.670	(0.9628)	-0.03290****	(0.0000)	0.069814	11.708	0.027	S4	7.14422	(0.08395)	-0.0212****	(0.0000)	0.05917
D6	-12.2343	(0.9234)	-0.01828****	(0.0000)	0.086784	6.903	0.016	S5	5.13531	(0.00557)	-0.01413****	(0.0000)	0.06946
D7	1.7990	(0.9784)	-0.01015****	(0.0000)	0.051260	3.382	0.008	S6	4.74986	(0.00018)	-0.01246****	(0.0000)	0.05722
D8	4.5845	(0.8822)	-0.02152****	(0.0000)	0.255672	1.168	0.003	S7	4.81154	(0.0000)	-0.01272****	(0.0000)	0.04685
S8	383.840****	(0.0000)	-0.00968****	(0.0000)	0.02327	2.558	0.006						
Original series	-131.169	(0.9941)	0.00675	(0.4292)	6.77×10^{-5}	432.474	1						

Table 4.6 Results of the test for leverage effect in daily data. Model tested, notations and significance indicators are similar to those specified in Table 4.5. p-values are reported in parentheses.

4.2.3.3.3 Robustness checks on the presence of leverage effect in intraday data

Why is there conflicting evidence for 5-minute smooths? From the previous analyses, it can be concluded that while the presence of a daily leverage effect is confirmed both by our graphical inspection and quantitative test, the implications for intraday data is mixed. That is, the link between the regression results regarding the 5-minute smooth series (in Table 4.5) and the CCF plots (in Figure 4.14) is not altogether strong. Specifically, panel B of this Figure suggests an apparent leverage effect across all levels for the smooth series and this effect seems to become stronger at lower frequencies. However, the quantitative test for 5-minute smooths rejects the existence of this effect at all but the lowest frequencies.

There is a simple explanation for the different interpretation of the plots and the tables. To our understanding, the quantitative test only deals with correlation at lag 1 rather than at greater lags (recall Equation 4.2.3). Because the CCFs of smooths (from panel B, Figure 4.14) yield either positive or insignificant coefficients at lag 1 for level 1 to 5, they are actually in line with panel B of Table 4.5. The same explanation goes for 5-minute details and daily data: in these cases the lag 1 coefficients are generally significant. So the real issue here is that our interpretation of the existence of leverage effect depends upon the lag at which we expect to observe this effect. Unfortunately, there is no clear answer whether it should be lag 1 (in which case we should use the table's results) or higher positive lags (when the plots are more intuitive). Therefore, we suggest taking complementary implications from both approaches, where possible.

More discussions on the CCF plots Although we have argued that the intuition behind the increasing significance of leverage effect, if it exists, may be related to the rational reaction of the market thanks to the cumulative information gathered at longer time horizons, we still have to be cautious when interpreting the implications of the CCF between wavelet smooths. The reason is that the variance of two time series decreases as they become smoother, therefore we would expect their cross correlation coefficients to become higher in magnitude. As such, it is possible that the increasing significant coefficients reported in Figure 4.14 are 'spurious' in the sense that they may merely reflect the artificial linear relation introduced by a smoothing operation. Furthermore, the significant correlation could merely be a result of fixed confidence interval. We know that the confidence interval is fixed because the sample size is maintained at all

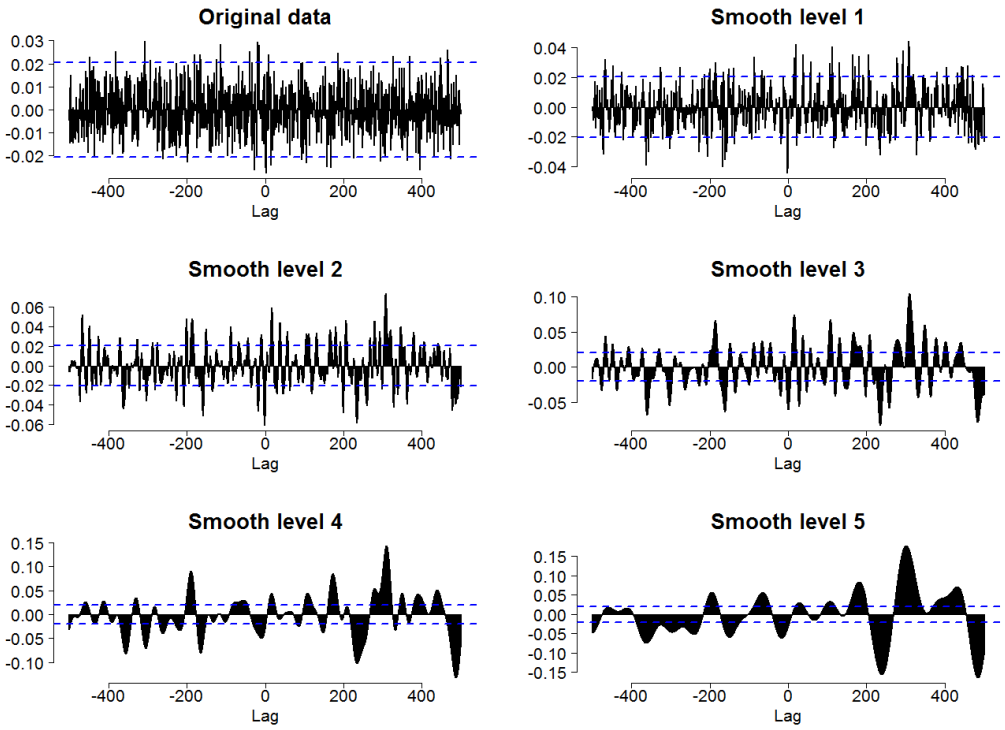


Figure 4.15 Multi-scale cross correlograms between two random independent variables. Blue dashed lines indicate the 95% confidence interval.

frequencies due to the MODWT transform [36]. However, the assumption of an asymptotically normally distributed cross-correlation estimator may no longer be appropriate, since the statistical properties of the smooth series may be different from those of the original signal.

To investigate the possibility of an artificial asymmetric correlogram pattern introduced by wavelet analysis, we repeat the procedure of plotting multi-scale CCFs with two independent, randomly generated time series which originally do not have any significant correlation. Figure 4.15 shows that, while the smoothing operation indeed strengthens the significance of the cross-correlation, it is difficult to spot any systematic evidence of asymmetric pattern across the levels. Therefore, we have more confidence when interpreting the implications of Figure 4.13 and 4.14. Nevertheless, although it might be true that these asymmetric correlograms are not spurious, the notion of no significant intraday leverage effect as implied by panel B, Table 4.5 still needs to be carefully considered.

An alternative approach to check for spurious leverage effect is to calculate cross-correlation between returns and volatility at intraday frequencies other than

[36] Refer to subsection A.2.3, Appendix A for further discussions.

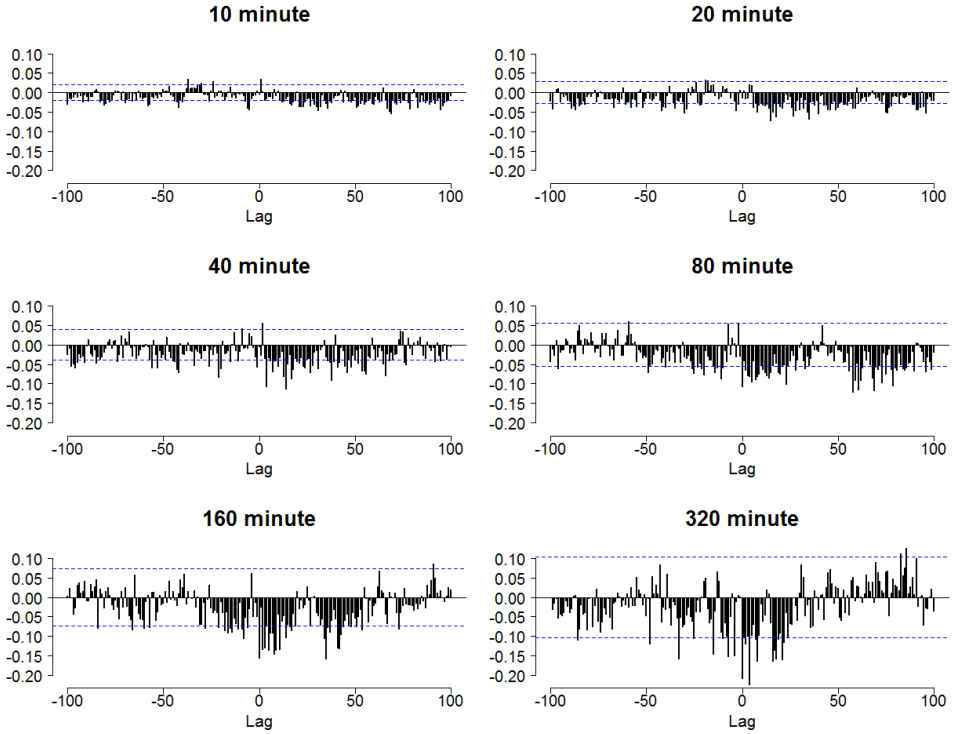


Figure 4.16 Cross-correlograms between intraday returns and volatility. Data constructed from decreasing sampling frequencies incorporate more informative, and thus are expected to exhibit stronger leverage effect. Blue dashed lines indicate the 95% confidence interval. Since the number of observations decreases (as the sampling frequency decreases), the confidence interval widens. Consequently, in general there is little significant cross-correlation. However, we can still see stronger negativity at positive lags (especially at the lags closer to zero).

5-minutes [37]. To do this, we compute intraday returns by re-constructing the volume weighted price series (recall the formula in subsection 4.1.2) with a trading frequencies corresponding to 10, 20, 40, 80, 160 and 320 minutes (in a way, these time series contain the information similar to that of the smoothed 5-minute data). The new data are expected to exhibit stronger leverage effect, since more information is accumulated at larger time scales (for example, a 20-minute return should incorporate cumulative information four times as much as that of a 5-minute return). However, as can be seen from Figure 4.16, there is little evidence of significant cross-correlation between returns and volatility when studying intraday frequencies other than 5-minute, although the asymmetric pattern becomes stronger at lower frequencies.

So the question of whether there exists a genuine leverage effect in intraday

[37] Another alternative test would be to compute the CCF using cumulative 5-minute returns instead of raw 5-minutes returns, as the leverage effect, if presents, would be more convincing since the returns now incorporate information accumulated over time. We thank Graeme Guthrie for pointing this out. Although we do not pursue this direction in the current paper, it could be included in future revisions.

data remains somewhat inconclusive. On the one hand, intraday data constructed from frequencies lower than 5-minute reveals little significant correlation. On the other hand, the multi-scale asymmetric correlation pattern observed in 5-minute data is not spurious. Nevertheless, we are inclined to believe the theory of no intraday leverage effect for Citigroup. After all, it is not unreasonable to assume that, for a highly liquid stock like Citigroup's, the information conveyed by intraday price changes may be too little (or maybe too noisy) to alter investors' perception about a company's level of risk. Analogously speaking, most traders would not bother searching for clues for their next trade from information arriving every 5-minutes. For thinly traded stock the situation is different in that rather than having too noisy information, there are not many intraday transactions; but the conclusion stays the same.

4.2.3.3.4 In retrospect and a reversed causality of the return-volatility dynamic

Up to this point we have discussed the negative contemporaneous relationship between returns shocks and unexpected change in volatility with explanations related to the effect of leverage. Now we explore what this relationship implies. Regarding the same problem, French et al. (1987) provide a summary of this phenomenon and argue that "*[a] positive unexpected change in volatility (and an upward revision in predictable volatility) increases future expected risk premiums and lowers current stock price.*" (p.27). Additionally, to test whether leverage effect solely explained this relationship, these authors tested for the proposition of Black (1976) and Christie (1982): should it be the case, the slope coefficient (or elasticity) from a regression of the percentage change in standard deviation on the percentage change in stock price must be between zero and -1. This regression is:

$$\ln(\sigma_{m,t}/\sigma_{m,t-1}) = \alpha + \beta \ln(1 + R_{m,t}) + \epsilon_t \tag{4.2.4}$$

where $R_{m,t}$ is monthly returns, $\sigma_{m,t}$ is the monthly volatility measure computed as standard deviation of daily returns within month t. We can see the analogy of this Equation and Equation 4.2.3, save for the percentage adjustment and the frequency difference. The fundamental common-ground is to compute and interpret the elasticity of the change in volatility with respect to the change in price levels. From previous results we know that our (significantly negative) β estimates satisfy the condition of being larger than -1 at all levels, thus supporting the sole contribution of leverage effect.

Though in favour of a "strong negative relation between the unpredictable

component of stock market volatility and excess holding period returns", the findings of Black (1976) and French et al. (1987) reject the hypothesis that leverage alone can explain this relationship. Subsequently, they simply interpret it as an indicator of the positive relationship between expected risk premium and (ex-ante) expected/predictable volatility. The latter relationship is very well documented in traditional financial literature and is an intuitive one (i.e. higher risk demands higher returns). Therefore, it is reasonable to assume our results point to the same conclusion: a negative relation in unexpected components of volatility-returns dynamic implies a positive relation in its expected components.

The lack of explanatory power of the leverage effect hypothesis, as pointed out in French et al. (1987), together with the evidence of time varying risk premiums (see also Campbell and Hentschel (1992)) motivate the second branch of theory: the volatility feedback effect hypothesis. Accordingly, if there is an anticipated increase in current volatility, a higher risk premium is expected (or higher returns are expected) resulting in an immediate decline in stock price (or a negative impact on ex-post returns) (Bekaert and Wu (2000)). In this school of thought, the volatility-return causal relationship proposed by the leverage effect is reversed. An important point is that this theory only holds when there exists a positive relation between the expected parts of returns and volatility.

Evidence of this effect is partially reflected via the long-memory behaviour of volatility. Intuitively, current volatility has a positive impact on expected returns and expected volatility. This in turn has a positive impact on the subsequent period's volatility. Therefore, we expect to see a correlation between present and past volatilities whose magnitude decreases as the time distance increases.

4.2.4 Multi-scale long-memory estimates

Empirical evidence (including ours) so far supports the existence of highly persistent conditional variance processes in financial data. This fact effectively motivated the development of the IGARCH model. Also, Vuorenmaa (2005) pointed out the crucial fact of time-varying long-memory parameter and suggested that volatility persistence is more likely to be the case in turbulent times. In this section, we shall examine the credibility of these arguments by investigating the long memory behaviour of the volatility process not only at different times, but also at different scales/frequencies.

4.2.4.1 Estimating with rescaled range analysis and related methods

The following table reports the Hurst exponent estimated via the various methods described in Chapter 2 and Appendix B for both daily returns and volatility series:

Method	R/S	aggVar	diffaggVar	AbaggVar	Per	Peng	Higuchi
r_{wins}	0.4639 (0.0501)	0.5279 (0.0293)	0.4700 (0.1123)	0.5264 (0.0480)	0.5199 (0.0099)	0.4791 (0.0193)	0.4914 (0.0426)
$\|r_{wins}\|$	0.738 (0.0620)	0.9214 (0.0452)	0.8737 (0.1390)	0.9192 (0.0305)	0.6317 (0.0095)	0.7120 (0.0965)	0.9678 (0.0312)

Table 4.7 Estimates of Hurst exponent with different methods suggested by Taqqu et al. (1995), for daily returns and volatility series. Sample range is from 03 Jan 1977 to 31 Jul 2013. Standard errors are reported in parentheses. Nomenclatures are as follows: (1) R/S : Rescaled range; (2) aggVar: Aggregated variance; (3) diffaggVar: differenced variance; (4) AbaggVar: Absolute moments; (5) Per: Periodogram; (6) Peng: regression of residuals; (7) Higuchi: Higuchi's method.

Note that exploring the relative performance of our estimating techniques is beyond the scope of this paper as the topic remains open, and subject to much ongoing research. Nevertheless, the methods applied yield comparable estimates. Moreover, according to the ranking of performance provided in Section 2.3, Chapter 2 which shows superiority of the wavelet-based estimator, the Peng method and Periodogram method seem to offer the next best estimators (recall that these methods take turns to occupy the 2nd and 3rd places in all simulated experiments). Therefore, wherever wavelet MLE is not used we shall refer to these two estimators as our primary benchmarks to examine long-memory behaviour.

For daily volatility process, overall we observe a fairly strong long run dependence: judging from the propositions introduced in Chapter 2, our estimates of Hurst index (ranging from 0.7120 to 0.9678) indicate a certain degree of predictability of Citigroup's volatility. On the contrary, the returns series exhibits only weak dependence: the corresponding estimates of the Hurst index are in close proximity to 0.5, which is the theoretical value characterizing a random walk, or a Brownian motion. It is important to note that since most Hurst index estimates for returns do not differ from 0.5 at any reasonable significance level, for our purposes it is sufficient to regard the returns series as white noise.

The evidence of dependence structure of both returns and volatility is further supported by the visual features of their respective autocorrelation functions as shown in Figure 4.5. In particular, the absolute returns ACF decays at a hyper-

bolic rate whilst there is little/insignificant serial correlation among returns.

As with previous experiments, we proceed to investigate the difference of our estimates throughout the scales. Specifically, we apply our estimating methods to the smooth series obtained from MRA of daily volatility at different levels. Maximum decomposition depth is set to 8. Estimated results are reported in Table 4.8. All estimates are significantly different from 0.5 at the 1% level.

Method	S1	S2	S3	S4	S5	S6	S7	S8
R/S	0.7379	0.7799	0.7978	0.8918	0.9413	0.9903	0.9983	0.9877
	(0.0493)	(0.0381)	(0.0415)	(0.0301)	(0.0281)	(0.0277)	(0.0217)	(0.0142)
aggVar	0.9229	0.9254	0.9325	0.9454	0.9624	0.9761	0.9872	0.9939
	(0.042)	(0.042)	(0.044)	(0.0470)	(0.0503)	(0.0526)	(0.0531)	(0.0516)
diffaggVar	0.8784	0.8806	0.9649	1.2305	1.5381	1.8263	2.1056	2.1717
	(0.1249)	(0.1237)	(0.1233)	(0.1032)	(0.1324)	(0.1668)	(0.2706)	(0.4552)
AbaggVar	0.9211	0.9235	0.9316	0.9461	0.9714	0.9858	0.9943	0.9992
	(0.0244)	(0.0229)	(0.0248)	(0.0278)	(0.0318)	(0.0341)	(0.0352)	(0.0345)
Per	0.8019	0.8296	1.7694	4.1678	5.584	5.6756	5.6259	5.6399
	(0.0223)	(0.0222)	(0.0470)	(0.1901)	(0.2135)	(0.2090)	(0.2003)	(0.1864)
Peng	0.7757	0.895	1.1136	1.3758	1.662	1.8296	1.9295	1.9836
	(0.0845)	(0.0753)	(0.0718)	(0.0915)	(0.1215)	(0.1277)	(0.1136)	(0.0944)
Higuchi	0.9696	0.9694	0.9694	0.9692	0.9691	0.9689	0.9686	0.9688
	(0.0294)	(0.0293)	(0.0293)	(0.0293)	(0.0293)	(0.0293)	(0.0293)	(0.0292)

Table 4.8 Hurst index estimates with level-specific smooth series obtained from MRAs of volatility. Standard errors are reported in parentheses. Decomposition depth is set at 8 levels. Nomenclatures are similar to those specified in Table 4.7.

We can see that in general, the long run dependence behaviour of volatility persists and becomes stronger at higher levels. Intuitively, as the series gets smoother, the trending behaviour becomes clearer and is picked up by our estimators. Interestingly, the Peng and Periodogram methods, as well as the Difference Aggregated Variance method all yield estimates exceeding 1 (from level 3 and 4), suggesting non-stationarity at higher levels; or that differencing is needed to achieve stationarity again. That said, one should not take too seriously the face values of the Hurst index estimates at very low frequencies. It appears that some of the methods are 'breaking down' when being applied to excessively smooth data, for example, the Periodogram exhibits Hurst exponent greater than 5.6. The Aggregated Variance, Absolute aggregated Variance and Higuchi methods yield relatively stable estimates and standard errors.

4.2.4.2 Estimating with wavelet-based maximum likelihood method

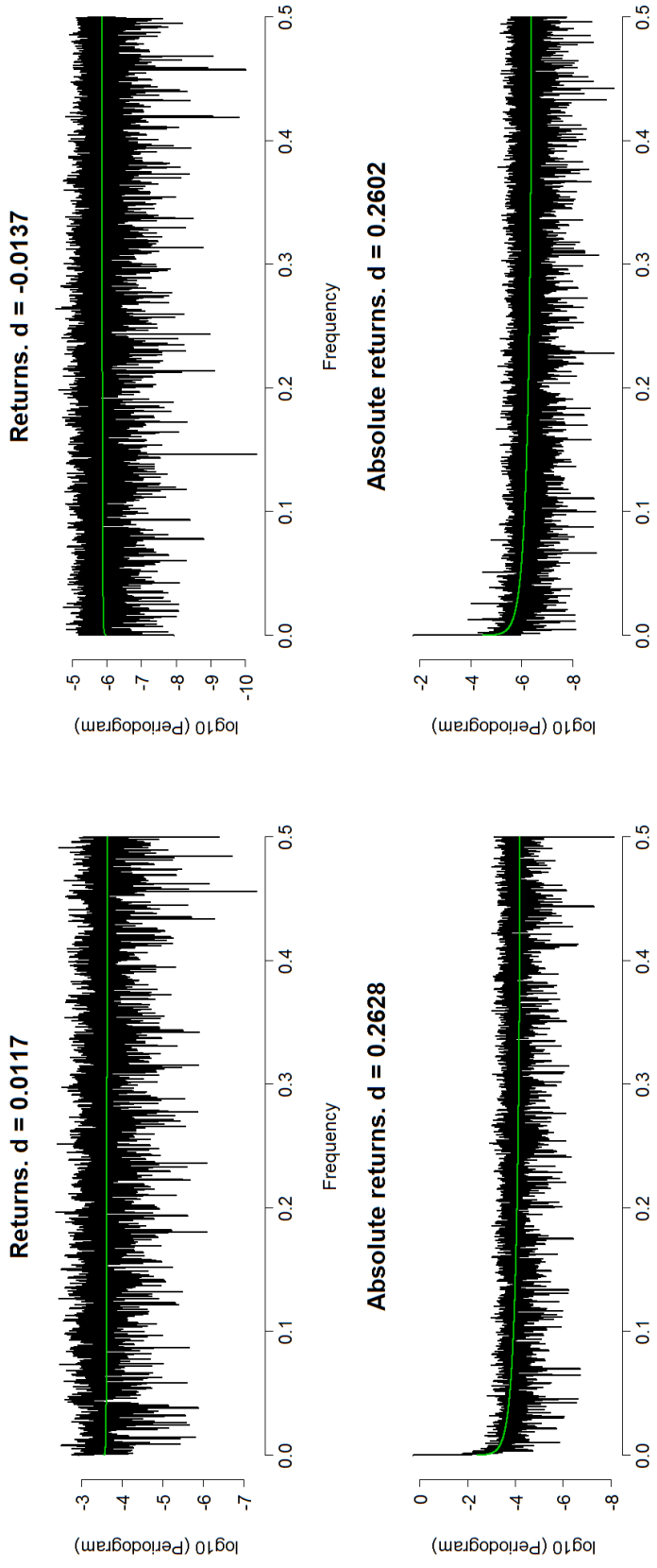
Following the procedure described in Section 2.2.4, Chapter 2, we now apply the wavelet-based maximum likelihood estimator to our daily data. We need dyadic length time series for this analysis, which is why we exclude the first 1036 observations in our sample. That is, our new sample now ranging from from 06 Feb 1981 to 31 Jul 2013 (for a total of $8192 = 2^{13}$ observations). Here we use an LA8 wavelet with decomposition depth set at 13 levels.

The estimate of d by wavelet-based MLE indicates a value of 0.0117 ($H = 0.5117$) for returns and 0.2628 ($H = 0.7628$) for volatility. Corresponding standard errors are 0.00047 and 0.00018 indicating significance at the 0.1% level. [38]. The SDF of returns series is effectively a horizontal line, analogous to that of a Brownian motion. Consistent with Gençay et al. (2002)'s findings, the proposed model seems to capture the long memory feature of the volatility process quite efficiently.

From Panel A, Figure 4.17 it can be observed that despite some random disturbances, overall the periodogram of the absolute daily returns closely tracks the hyperbolically decaying feature of the spectrum of a fractionally integrated process with a fractional difference parameter $d = 0.2628$. Ostensibly, this estimate is not directly comparable to that of the previously discussed methods, as our sample range has been reduced to suit the dyadic condition. Nevertheless, if the volatility process is truly 'self-similar', reducing the data range should not affect its long-memory behaviour across scales. Thus our estimate is sufficient to give an idea of how strongly volatility persists.

The same method is applied to the intraday sample, yielding $d = -0.0137$ for returns and $d = 0.2602$ for volatility. Again we observe long-memory in the latter signal while the returns series is indistinguishable from a white noise process. More striking still, is the similarity between the estimates of d for daily and 5-minute data. This shows how long-memory behaviour is preserved at different frequencies thanks to the self-similarity feature of the underlying process.

[38] The formula to derive the model standard errors can be found in subsection 2.2.4, Chapter 2 or more thoroughly in Gençay et al. (2002), p.172-174.



(a) Panel A: daily series

(b) Panel B: 5-minute series

Figure 4.17 Spectral density function (on a $\log_{10}$ scale) of the daily (panel A) and 5-minute (panel B) returns and volatility processes of Citigroup. Highest frequency being 0.5 corresponds to the Nyquist frequency. In all plots the green line indicates theoretical spectrum of long-memory process with values of d estimated by the waveMLE method.

When looking more closely at the SDF of 5-minute volatility series, we notice that there are 4 small 'spikes' at some frequencies close to zero. Further examination by increasing the scale of x-axis reveals that these 'local maxima' correspond to several periodic components. With regards to the spectral methodology discussed in Appendix A these frequencies are called 'prominent' frequencies, meaning the periodic components corresponding to them contribute the most to the total variance (energy) or the signal. Likewise, the sum of the (stationary) sine waves fluctuating at these frequencies should resemble the underlying 'trend' of the signal closely.

As shown in the bottom plot of Figure 4.18, these frequencies are 0.0128, 0.0256, 0.0385 and 0.0512, which respectively correspond to periods 388.56, 195.06, 129.86, 97.53 (in units of minutes). We do not observe any other prominent frequencies in the form of notable local maxima (spikes). Not surprisingly, of these four local peaks, the highest value is observed at frequency $f = 0.0128$ whose period represents the daily seasonality (as we have 390 minutes in a trading day, 388.56 minutes approximates one day). What is interesting is the other three frequencies represent periodicities associated with half a day, a third of a day and a quarter of a day, respectively. This suggests that there are more subtle periodic pattern exists within the dynamic of Citigroup volatility structure, other than the daily periodicity discussed. This finding could be an opening for a future research, which might investigate, for example, whether the transition between morning and afternoon trading sessions could explain why volatility peaks at the middle of a trading day.

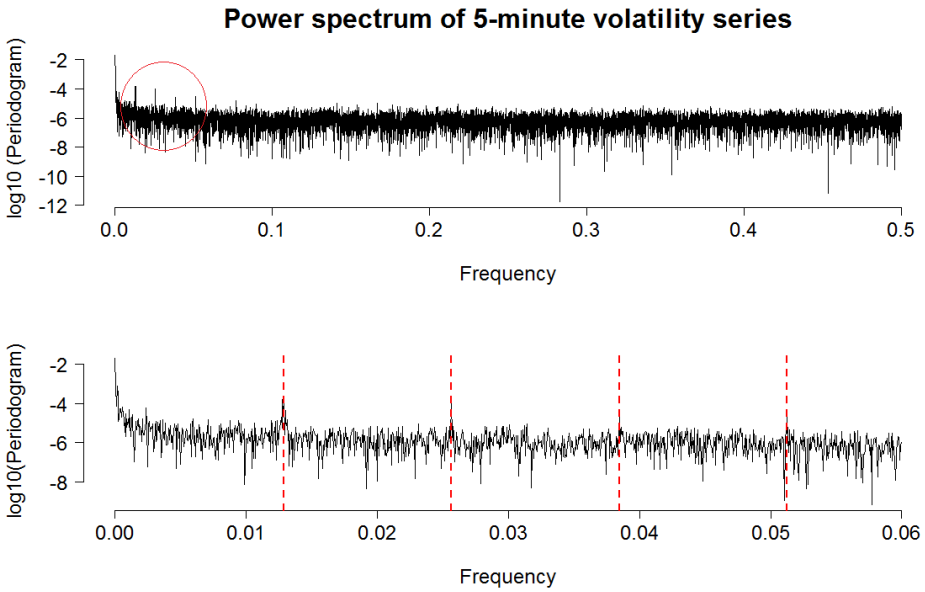


Figure 4.18 Spectral density function of 5-minute volatility. The red circle delineates the four spikes at smaller scale plot. The red dashed lines indicate the prominent frequencies corresponding to the periodic components that contribute the most to total energy of original signal/process.

4.2.5 The relevance of time-varying Hurst index estimates in returns predictability

This section 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 arguments may only be true from an aggregated perspective. As Mitra (2012) pointed out, it is not unusual to observe deviation from market efficiency in the form of Hurst exponent 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 Dow-Jones Industrial Average index (DJIA) exhibits remarkable heterogeneity in its dependence structure over time: most notably are the strong persistence during the period 1930-1980 and mean-reversion during the period 1980-2004.

4.2.5.1 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 in a very 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 efficient market. One straightforward way to do this is to perform a simplified Monte Carlo simulation [39] 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. Based on these calculations, our new benchmark for 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% confidence to state that the true population value of Hurst index of a length $n = 1,024$ random time series lies within the interval $[0.4839, 0.6378]$.

Consequently, when examining the time-varying nature of Hurst exponent, we can relatively 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 M. Hull and McGroarty (2013).

Next, to compute our time-varying estimate series we follow Qian and Rasheed (2004)'s approach, i.e. we adopt a 'rolling' estimate of Hurst index. First, the index is estimated for the first time window of 1,024 days (approximately 4 years) [40] using the Rescaled range method, since it is arguably the most popularly used method. 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

[39] 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 (Guthrie (2013)).

[40] According to Qian and Rasheed (2007), one of the reasons for the choice of a length $n = 1,024$ rolling window is that Peters (1994) had documented a four-years cycle exhibited by the DJIA index series. Although we only examine Citigroup stock price here, the cycle is relevant since Citigroup was one of DJIA's main components until recently.

1,024 observations, for a total of 8,204 observations starting from the beginning of 1981. As can be seen from Figure 4.19, the estimates for Hurst 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 8,204 observations we observe weak anti-persistence (mean reversion) behaviour (58.31%) and weak persistence behaviour (40.22%) (although there is some evidence of strong anti-persistence in mid 1998 and strong persistence some time from 2008 to 2010).

In addition, there are two very notable twists 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 recent serious crises in the finance world. Relating these events to the corresponding movements of the stock price (in Figure 4.2, subsection 4.1.1.2.1) 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) 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 other period.

Another possible explanation for the strong mean reverting behaviour observed in 1997-1999 would be the impact of 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 from the one in 1997, as it reflects the realization of a fundamental weakness of the U.S. financial system, rather than a short term over-reaction, thus explaining the sharp increase (rather than decrease) of the Hurst index from 2007 when market confidence continued to fall and price changes exhibited long-memory. Overall, a time-varying Hurst index is consistent with the notion that market efficiency should always be considered from a dynamic, evolutionary viewpoint (recall our discussion on the "Adaptive Market Hypothesis" in Chapter 2).

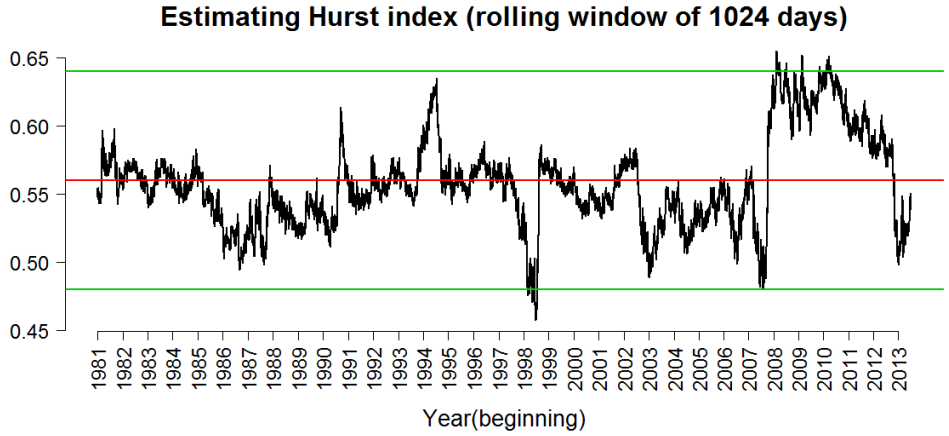


Figure 4.19 Time-varying Hurst index estimate of daily returns of Citigroup, Inc. Using R/S method, Hurst index is computed for a rolling window of 1,024 days, starting from 03 Jan 1977. The first observation is computed at the beginning of 1981. The red line indicates the efficient market benchmark for the Hurst index estimate ($H = 0.56$), whilst the green lines approximate the 'upper' and 'lower' bounds of the 95% confidence interval ($H = 0.48$ and $H = 0.64$, respectively).

4.2.5.2 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). In its simplest form the (equally weighted) SMA is 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:

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

We can see that the daily price series is just a special case of SMA with $n = 1$. Without any doubt these averaging metrics are among the oldest and most popularly used indicators of existing trends 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, Lakonishok, and LeBaron (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 purpose, we examine 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 [41]. 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' [42].

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 start with adjusted closing prices instead. To illustrate, we plot four particular SMAs series (which are commonly used in practice) along side Citigroup's daily adjusted closing price for the year 2007 in Figure 4.20. 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 increases with the length of the time window on which the SMA is based. In general, a shorter-term average system 'moves' faster and creates 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

[41] 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.

[42] An alternative strategy may 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 moving average 'cross-over' (MACO). Another iteration 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 rule are robust in the sense that they are strong indicators of an upward/downward trend, as can be seen in Figure 4.20, 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 towards 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.

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.

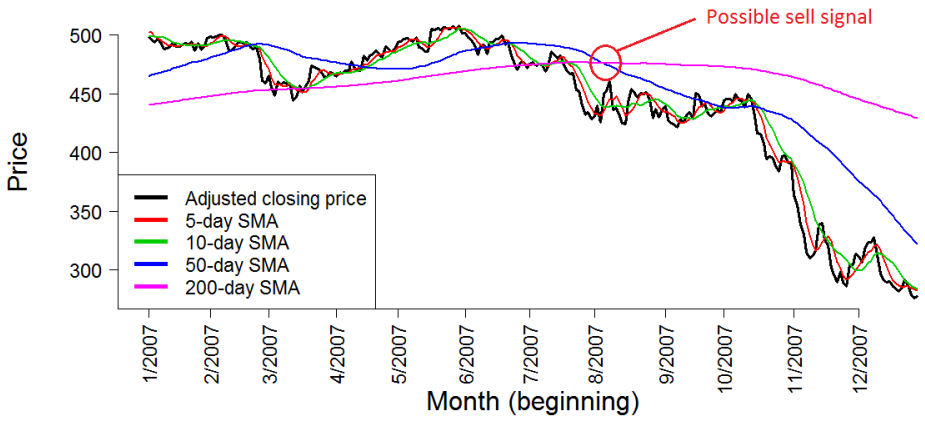


Figure 4.20 Plot of adjusted closing price for the year 2007, together with 5-day, 10-day, 50-day and 200-day SMA. To maintain the number of observations the SMAs are computed from a sample longer than the year 2007, since all SMAs are lagged series. The red circle indicates a strong sell signal observed when the 50-day SMA 'crosses over' the 200-day SMA.

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 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 taking Qian and Rasheed (2004)'s approach. That is, our time-varying daily Hurst index estimates are based on a rolling window of 1,024 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}) calculated from section 4.1.1. For example, if the original return is -0.02 on a specific day then at the end of the day we would realize a returns of 0.02 (or -0.02) should we receive a sell (or

buy) signal on that day. 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 1,024 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 4.9 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 interval 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 4.10, 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.

Conclusion When analysing the time-varying nature of the Hurst index series and its correlation with trading returns, we noted 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? And more radically, what introduce these instances of inefficiency?), 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 length chosen (1,024 days) are merely illustrative and should be taken as indicative only. It is possible that with other specifications we shall obtain different interpretations. The implications from trading techniques based on other indicators (such as the Exponential Weighted Moving Average (EWMA)) could also be important. In addition, the question of whether mean-reverting trading strategies perform better when the series exhibit mean-reversion behaviour remains open. Undoubtedly, for possible future research, a more complete survey is in demand to provide a better understanding of these intriguing observations.

Our main finding in this ending section 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). Furthermore, this author cautiously argues that evidence of price predictability does not necessarily imply market inefficiency, and that such conclusion can only be drawn if the trading rule consistently outperforms some passive strategies, e.g. a risk-free investment. Additionally, Bessembinder and Chan (1998), among many others, agree that the forecast value of technical analysis may not provide the ability to "beat the market" (as the actual trading costs may dissolve all gross profit) and in that case 'abnormal profit' is by no means inconsistent with efficiency, but merely a result of randomness.

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 4.9 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 4.10 Average trading returns corresponding to different values of the Hurst index estimates.

Chapter 5

Concluding remarks

In this Chapter we first conclude the paper with an overview of the main findings and some of their direct implications, then analyse potential venues for future research.

5.1 Summary of the main findings

Our first contribution is the applications of various wavelet methodologies in studying volatility dependence structure, specifically in conjunction with conventional techniques such as GARCH models and Hurst index estimate methods. This type of combination is, to our understanding, relatively innovative. As illustrated in Appendix A, wavelet decomposition technique provides us with a powerful tool to break down the signal of interest into multiple frequency-based components. Compared to the traditional time series and spectral analyses, this is a significant improvement as wavelet operations not only preserve time-domain information, but also capture the evolution of signals across different time scales. In this way we combine the strengths of the two branches of methodology, while not being limited by their respective weaknesses. Ultimately, the wavelet method helps reveal more complete insights into the implications of the time-varying volatility dependence dynamic:

- When fitting a GARCH model, we obtained a better goodness-of-fit with low frequency data, most notably at frequencies associated with trading periods longer than 2 months. This indicates the importance of long-term components of the volatility process. To some extent, the role of low frequencies is related to the role of long-term business cycles and large financial institutions. All in all, investors with long trading horizons seem to be more important in determining market dynamics, than is currently considered to be the case in the literature.

- The asymmetric relationship between returns and volatility, as modelled by various GARCH specifications, is a direct indicator of the so-called leverage effect, which explains the rise of volatility following a fall in stock price

(and possibly increased leverage) due to a change in the firm's level of risk. Wavelet analysis shows that while the effect seems to be more robust at lower frequencies when studying daily data, examination of such effect in intraday data gives mixed results. In general, there is not enough evidence to confirm the presence of this effect at intraday frequencies.
Intuitively, the magnitude of the leverage effect depends on the time necessary for information about the firm's performance to be accumulated and reflected in its stock price. Therefore, the longer the time horizon (or the larger the scale), the greater the impact a current shock in returns would have on future volatility, and the stronger the leverage effect becomes. As such, changes in stock price within a time frame of one day may not be enough to affect investors' perception of the company's risk.

- In terms of estimating the Hurst index, we find that in many cases, the wavelet-based estimator developed by Gençay et al. (2002) outperforms various well-established estimators that have been extensively studied in the literature.

Our second contribution is expressed via the estimation of long-memory parameter of financial time series: in line with previous researches, in general we find a long-run dependency in volatility with the Hurst index estimated at approximately 0.7; while the returns series does not exhibit such behaviour, with a Hurst index 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, adopting 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 with a rolling window of 4 years varies widely, being higher whenever returns' dependence increases.

Furthermore, this time-varying Hurst index series is positively correlated with the profits 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 capitalize on the 'trends' of 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 improving their strategies. For example, traders may opt for trend-following rules (or alternatively, mean-reverting rules) when stocks exhibiting Hurst index higher (lower) than 0.5.

5.2 Caveats and possible openings for future research

With respect to our main findings, the paper still exhibits a number of non-trivial caveats which can be listed as follows:

- The application of GARCH models in Chapter 4 is subject to an inherent weakness: the input (or the information set used to estimate model parameters) and the ex-post volatility measure are both related to returns. This introduces a sort of 'circular reasoning' problem when we compare the fitted volatility estimates and the ex-post measure. That being said, this is the standard approach which has been extensively adopted in many theoretical and empirical studies. In any case, it would be more convincing if we used a realized volatility measure constructed from high-frequency data, as illustrated in Chapter 3. Because of data and time limitation, however, we were unable to build a sufficiently long realized volatility series to use in the studies in Chapter 4. With access to the database of Bloomberg (hosted at Victoria University of Wellington) hopefully we can improve our study in the near future.

  Furthermore, although we have used 5-minute returns to construct a daily realized measure of volatility for assessing GARCH estimates, we have not explicitly modelled 5-minute data with GARCH. It may be that the intraday data provide a better goodness-of-fit which can further confirm the self-similarity of the data generating process.

- Additionally, there is the problem with determining the appropriate data sampling rate, which is a matter of balance between preserving information (reducing estimator's bias) and mitigating the microstructure effects (reducing noise) at high frequencies. Although 5-minute intervals seem to be a good candidate as discussed earlier, the choice is arbitrary at best and is currently a subject of much debate among academics.

- Although most of our evidence points to the importance of low-frequencies component of the latent volatility factor, we have not tackled the intuition of these findings to a satisfying extent. This requires better understanding of economic and business cycles. Specifically, when saying that trading activities associated with periods longer than 2 months play a crucial role in determining the volatility data generating process, we have not answered the question as to why this might be. Also, we need to verify if such activi-

ties can really be attributable to investors with long-term trading horizons. Indeed, without further examination of these problems our findings would be of little use to practitioners, except for deriving some general advice when building their strategies.

- In a recent paper, Kim and Jung (2013) point out that the margin requirements, which are imposed by stock market regulators as a tool to moderate the level of volatility, should be adopted with more caution than was suggested before. In particular, when examining the conventional assumption that increasing margin level helps reduce volatility, they argue that this negative relationship might only be true for long run volatility and not for short run volatility. Specifically, higher margin level impacts the market in two contradictory ways, which depend on two conditions known as the "liquidity effects" and "speculative effect": First, margins discourage stabilizing trades and force investors to exit. The lack of liquidity then increases volatility in the short term; second, higher margins reduce traders' speculation activities, thus reducing volatility in the long term.
  With regards to our study, we note that the volatility process may be dominated by long-term (rational) traders, and it will be fruitful to see how an increased margin policy affects the level of varying volatility components, both short-run (high frequency) and long-run (low frequency) in this context.

- As pointed out at the end of Chapter 4, the choice of trading strategies and window length of the SMAs are only indicative. It is arguable that such strategies happen to be exclusively correlated with heavily traded stock or stock that exhibits long-memory. The possibility of whether other strategies (such as the one based on mean-reversion) or different signal indicators (such as the EWMAs) could yield better insights remains open. Additionally, it would be more practical if we could somehow incorporate the information conveyed by the Hurst index estimates to perform some sort of prediction of future market movements.

5.3 The bottom line

Our study, for all intents and purposes, aims 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.

Preferably, to extend our paper's implications to a wider group, in the future, we could conduct event studies to focus on specific 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. Alternatively, we could organize similar investigations on various indexes, from both developed and emerging markets, to draw more general conclusions.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723.

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance*, *51*(1), 169-204.

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

Andersen, T. G., & Bollerslev, T. (1997b). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, *4*(2), 115-158.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*(4), 885-905.

Andersen, T. G., Bollerslev, T., & Cai, J. (2000). Intraday and interday volatility in the Japanese stock market. *Journal of International Financial Markets, Institutions and Money*, *10*(2), 107-130.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (1999). *(Understanding, Optimizing, Using and Forecasting) Realized Volatility and Correlation* (Working Paper Series No. 99-061). New York University, Leonard N. Stern School of Business. Retrieved from `http://ideas.repec.org/p/fth/nystfi/99-061.html`

Anderson, R. M., Eomb, K. S., Hahn, S. B., & Park, J.-H. (2012). Sources of stock return autocorrelation [Working paper]. Retrieved from `http://emlab.berkeley.edu/~anderson/Sources-042212.pdf`

Atchison, M. D., Butler, K. C., & Simonds, R. R. (1987). Non-synchronous security trading and market index autocorrelation. *Journal of Finance*, *42*(1), 111-18.

Awartani, B. M. A., & Corradi, V. (2005). Predicting the volatility of the S&P500 stock index via GARCH models: the role of asymmetries. *International Journal of Forecasting*, *21*(1), 167-183.

Baillie, R., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*(1), 3-30.

Baqaee, D. (2010). Using wavelets to measure core inflation: The case of New

Zealand. *North American Journal of Economics and Finance*, *21*(3), 241-255.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *64*(2), 253-280.

Barunik, J., & Vacha, L. (2012a). Realized wavelet-based estimation of integrated variance and jumps in the presence of noise. *ArXiv e-prints*. Retrieved from `http://arxiv.org/abs/1202.1854`

Barunik, J., & Vacha, L. (2012b). Realized wavelet jump-GARCH model: Can time-frequency decomposition of volatility improve its forecasting? *ArXiv e-prints*. Retrieved from `http://arxiv-web3.library.cornell.edu/abs/1204.1452v2`

Bekaert, G., & Wu, G. (2000). Asymmetric volatility and risk in equity markets. *Review of Financial Studies*, *13*(1), 1-42.

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

Bessembinder, H., & Chan, K. (1998). Market efficiency and the returns to technical analysis. *Financial Management*, *27*(2), 5-17.

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section*, 177–181.

Blair, B., Poon, S.-H., & Taylor, S. J. (2002). Asymmetric and crash effects in stock volatility for the S&P100 index and its constituents. *Applied financial economics*, *12*(5), 319-329.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, *31*, 307-327.

Bollerslev, T., Cai, J., & Song, F. M. (2000). Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the US Treasury bond market. *Journal of Empirical Finance*, *7*(1), 37-55.

Bollerslev, T., & Jubinski, D. (1999). Equity trading volume and volatility: Latent information arrivals and common long-run dependencies. *Journal of Business and Economic Statistics*, *17*(1), 9-21.

Bollerslev, T., & Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, *73*(1), 151-184.

Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, *11*(2), 143–172.

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

Borland, L., Bouchaud, J.-P., Muzy, J.-F., & Zumbach, G. (2005). *The Dynamics of Financial Markets: Mandelbrot's multifractal cascades, and beyond* (Working paper No. 500061). Science & Finance, Capital Fund Management. Retrieved from `http://ideas.repec.org/p/sfi/sfiwpa/500061.html`

Brock, W., Lakonishok, J., & 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, U.K: Cambridge University Press.

Campbell, J. Y., & Hentschel, L. (1992, June). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, *31*(3), 281-318.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton, N.J: Princeton University Press.

Cannon, M. J., Percival, D. B., Caccia, D. C., Raymond, G. M., & Bassingthwaighte, J. B. (1997). Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series. *Physica A: Statistical Mechanics and its Applications*, *241*(3), 606-626.

Carmona, P., & Coutin, L. (1998). Fractional Brownian motion and the Markov property. *Electronic Communications in Probability*, *3*(12). Retrieved from `http://ecp.ejpecp.org/article/view/998/1211`

Cavalcante, J., & 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., Cleveland, W., Kleiner, B., & Tukey, P. (1983). *Graphical methods for data analysis*. Boston, MA: Duxury: Wadsworth International Group.

Chan, K. (1993). Imperfect information and cross-autocorrelation among stock prices. *Journal of Finance*, *48*(4), 1211-1230.

Chan, K., Chan, K. C., & Karolyi, G. A. (1991). Intraday volatility in the stock index and stock index futures markets. *The Review of Financial Studies (1986-1998)*, *4*(4), 657.

Chang, E. C., Jain, P. C., & Locke, P. R. (1995). S&P500 index futures volatility and price changes around the New York stock exchange close. *Journal of Business*, *68*(1), 61-84.

Cheng, P., Roberts, L., & Wu, E. (2013). *A wavelet analysis of returns on heavily traded stocks during times of stress* [Working paper]. (School of Economics and Finance, Victoria University of Wellington: Wellington, New Zealand.)

Chiang, A. C. (1984). *Fundamental Methods of Mathematical Economics*. New York: McGraw-Hill.

Chou, R. Y. (1988). Volatility persistence and stock valuations: some empirical evidence using GARCH. *Journal of applied econometrics*, *3*(4), 279-294.

Chow, G. C. (1983). *Econometrics*. McGraw-Hill.

Christensen, B. J., & Nielsen, M. (2007). The effect of long memory in volatility on stock market fluctuations. *The Review of Economics and Statistics*, *89*(4), 684-700.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, *10*(4), 407-432.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*(4), 841-862.

Clark, K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, *41*(1), 135-155.

Cohen, K. J., Hawawini, G. A., Maier, S. F., Schwartz, R. A., & Whitcomb, D. K. (1980). Implications of microstructure theory for empirical research on stock price behavior. *Journal of Finance*, *35*(2), pp. 249-257.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, *19*(90), 297-301.

Crowley, P. M. (2007). A guide to wavelets for economists. *Journal of Economic Surveys*, *21*(2), 207-267.

Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis With Applications in R*. SpringerLink (Online service). `http://link.springer.com/book/10.1007/978-0-387-75959-3/page/1`.

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

Dacorogna, M. M., Muller, U. A., Nagler, R. J., Olsen, R. B., & Pictet, O. V. (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance*, *12*(4), 413-438.

Daubechies, I. (1992). *Ten Lectures on Wavelets* (1st ed.). Society for Industrial and Applied Mathematics.

Dennis, B. (2010, Dec 7). Government sells remaining shares in Citigroup; investment to net $12 billion total profit for taxpayers. *The Wash-*

ington Post. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2010/12/07/AR2010120700091.html

Dickey, D. A., & 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.

Diebold, F. X. (1988). *Empirical Modeling of Exchange Rate Dynamics*. New York: Springer-Verlag.

Dieker, A. B., & 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. (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, *73*(1), 185-215.

Drost, F. C., & Nijman, T. E. (1993). Temporal aggregation of GARCH processes. *Econometrica*, *61*(4), 909-27.

Duffee, G. R. (1995). Stock returns and volatility: A firm-level analysis. *Journal of Financial Economics*, *37*, 399–420.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987-1007.

Engle, R. F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *The Journal of Economic Perspectives*, *15*(4), 157-168.

Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, *5*(1), 1-50.

Engle, R. F., Lilien, D. M., & Robins, R. P. (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica*, *55*(2), 391-407.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, *48*(5), 1749-78.

Engle, R. F., & 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.

Fleming, J., & Kirby, C. (2011). Long memory in volatility and trading volume. *Journal of Banking and Finance*, *35*(7), 1714-1726.

French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 3–30.

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

Gençay, R., Gradojevic, N., Selçuk, F., & Whitcher, B. (2010). Asymmetry of information flow between volatilities across time scales. *Quantitative Finance*, *10*(8), 895-915.

Gençay, R., Selçuk, F., & Whitcher, B. (2002). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. San Diego, Calif: Academic Press.

Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, *4*(4), 221-238.

Ghalanos, A. (2013). rugarch: Univariate GARCH models [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=rugarch` (R package version 1.2.2)

Gherman, M., Terebes, R., & Borda, M. (2012). Time series analysis using wavelets and GJR-GARCH models. *Signal Processing Conference (EUSIPCO)*, 2138-2142.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, *48*(5), 1779-1801.

Goupillaud, P., Grossmann, A., & Morlet, J. (1984). Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, *23*(1), 85-102.

Granger, C. W. J. (1966, Jan). The typical spectral shape of an economic variable. *Econometrica*, *34*(1), 150-161.

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

Granger, C. W. J., Spear, S., & Ding, Z. (2000). Stylized facts on the temporal and distributional properties of absolute returns: an update. *Statistics and Finance: an Interface*, 97-120.

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

Guthrie, G. (2013). *Monte Carlo simulation* [Lecture note]. (Course: FINA403/Derivative Securities. School of Economics and Finance, Victoria University of Welllington: Wellington, New Zealand.)

Haar, A. (1910). Zur theorie der orthogonalen funktionensysteme: Erste mitteilung. *Mathematische Annalen*, *69*(3), 331-371.

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

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models:

does anything beat a GARCH (1,1)? *Journal of Applied Econometrics*, *20*(7), 873-889.

Harris, L. (1986). Cross-security tests of the mixture of distributions hypothesis. *Journal of Financial and Quantitative Analysis*, *21*(1), 39-46.

Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D Nonlinear Phenomena*, *31*, 277-283.

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

Hull, J. (2006). *Options, Futures, and Other Derivatives*. Upper Saddle River, N.J: Pearson Education.

Hull, J., & White, A. (1996). *Hull-White on Derivatives: a Compilation of Articles*. London: Risk Publications.

Hull, M., & 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.

In, F., & Kim, S. (2005). The relationship between stock returns and inflation: new evidence from wavelet analysis. *Journal of Empirical Finance*, *12*(3), 435-444.

In, F., & Kim, S. (2013). *An Introduction to Wavelet Theory in Finance: a Wavelet Multiscale Approach*. Hackensack, NJ: World Scientific Pub.

In, F., Kim, S., & Gençay, R. (2011). Investment horizon effect on asset allocation between value and growth strategies. *Economic Modelling*, *28*(4), 1489-1497.

Jensen, M. J. (2000). An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *Journal of Economic Dynamics and Control*, *24*(3), 361-387.

Jensen, M. J., & Whitcher, B. (2000). *Time-varying long-memory in volatility: Detection and estimation with wavelets* [Working paper]. Retrieved from `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.4242`

Jorion, P. (1995). Predicting volatility in the foreign exchange market. *Journal of Finance*, *50*(2), 507-528.

Karatzas, I., & Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.

Kim, S., & In, F. (2010). Portfolio allocation and the investment horizon: a multiscaling approach. *Quantitative Finance*, *10*(4), 443-453.

Kim, S., & Jung, T. (2013). The effect of initial margin on long-run and short-run

volatilities in Japan. *Journal of East Asian Economic Integration*, *17*(3), 311-332.

Laurent, S., & Peters, J.-P. (2002). G@RCH 2.2: An Ox package for estimating and forecasting various ARCH models. *Journal of Economic Surveys*(16), 447-485.

Lee, J., Kim, T. S., & Lee, H. K. (2011). Return-volatility relationship in high frequency data: Multiscale horizon dependency. *Studies in Nonlinear Dynamics and Econometrics*, *15*(1), 7.

Liang, J., & Parks, T. W. (1996). A translation-invariant wavelet representation algorithm with applications. *IEEE Transactions on Signal Processing*, *44*(2), 225-232.

Ljung, G. M., & Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, *65*(2), 297-303.

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., & MacKinlay, C. A. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, *45*(1), 181-211.

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

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

Mallat, S. G. (2009). *A Wavelet Tour of Signal Processing: the Sparse Way*. Amsterdam: Elsevier /Academic Press.

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

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394-419.

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* (Vol. 1, p. 259-290). National Bureau of Economic Research, Inc.

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

Mandelbrot, B., & 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.

Martens, M. (2001). Forecasting daily exchange rate volatility using intraday returns. *Journal of International Money and Finance*, *20*(1), 1-23.

Martin, M. (1998, April 7). Citicorp and Travelers plan to merge in record $70 billion deal : A new No. 1: financial giants unite. *The New York Times*. Retrieved from http://www.nytimes.com/1998/04/07/news/07iht-citi.t.html

Masset, P. (2008). *Analysis of financial time series using Fourier and wavelet methods* [Working paper]. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1289420 (University of Fribourg (Switzerland), Falcuty of Economics and Social Science.)

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

McLeod, B., & Hipel, K. (1978). Preservation of the rescaled adjusted range. *Water Resources Research*, *14*, 491-518.

Melino, A. (1991). *Estimation of continuous-time models in finance* [Working paper]. (Department of Economics and Institute for Policy Analysis, University of Toronto.)

Merton, R. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, *8*(4), 323-361.

Mincer, J. A., & Zarnowitz, V. (1969). *The Evaluation of Economic Forecasts*. National Bureau of Economic Research, Inc.

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

Morgenson, G. (2011, Dec 3). Secrets of the bailout, now told. *The New York times*. Retrieved from http://www.nytimes.com/2011/12/04/business/secrets-of-the-bailout-now-revealed.html?pagewanted=all&_r=0

Muller, U. A., Dacorogna, M. M., Dave, R. D., Olsen, R. B., Pictet, O., & Von Weizsacker, J. E. (1997). Volatilities of different time resolutions - analyzing the dynamics of market components. *Journal of Empirical Finance*, *4*(2), 213-239.

Muller, U. A., Dacorogna, M. M., Dave, R. D., Pictet, O. V., Olsen, R. B., & Ward, J. R. (1995). Fractals and intrinsic time: a challenge to econometricians. In *Discussion paper presented at the 1993 international conference of the applied econometrics association.* Retrieved from http://ideas.repec.org/p/wop/olaswp/_009.html

Muller, U. A., Dacorogna, M. M., Olsen, R. B., Pictet, O. V., Schwarz, M., & Morgenegg, C. (1990). Statistical study of foreign exchange rates, empirical evidence of a price change scaling law and intraday analysis. *Journal of Banking and Finance*, *14*(6), 1189-1208.

Nason, G. P. (2008). *Wavelet Methods in Statistics With R*. SpringerLink (Online service). `http://link.springer.com/book/10.1007/978-0-387-75961-6/page/1`.

Nason, G. P., & Silverman, B. W. (1995). *The Stationary Wavelet Transform and some Statistical Applications*. Springer-Verlag.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH (1,1) model. *Econometric Theory*, *6*(3), 318-334.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347-70.

Nelson, D. B. (1992). Filtering and forecasting with misspecified ARCH models . *Journal of Econometrics*, *52*(1), 61-90.

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

Pagan, A. R., & Schwert, G. W. (1990a). Alternative models for conditional stock volatility. *Journal of Econometrics*, *45*(1), 267-290.

Pagan, A. R., & Schwert, G. W. (1990b). Testing for covariance stationarity in stock market data. *Economics letters*, *33*(2), 165-170.

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

Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical review. E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *49*(2), 1685-1689.

Percival, D. B., & Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge, England: Cambridge University Press.

Pesquet, J., Krim, H., & Carfantan, H. (1996). Time-invariant orthonormal wavelet representations. *IEEE Transactions on Signal Processing*, *44*(8), 1964-1970.

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

Peters, E. (1996). *Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility* (2nd ed.). New York: John Wiley & Sons.

Poon, S.-H., & Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, *41*(2), 478-539.

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

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

R Core Team. (2013). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rader, C., & Brenner, N. (1976). A new principle for fast Fourier transformation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *24*(3), 264-266.

Reider, R. (2009). *Volatility forecasting I: GARCH models* [Lecture note]. Courant Institute of Mathematical Sciences , New York University: New York, U.S. Retrieved from `http://cims.nyu.edu/~almgren/timeseries/Vol_Forecast1.pdf`

Resnick, S. (2007). *Extreme Values, Regular Variation, and Point Processes*. New York: Springer.

Richardson, M., & Smith, T. (1994). A direct test of the mixture of distributions hypothesis: Measuring the daily flow of information. *Journal of Financial and Quantitative Analysis*, *29*(1), 101-116.

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

Schwert, G. W. (1989). Why does stock market volatility change over time? *The Journal of Finance*, *44*(5), 1115-1153.

Taqqu, M., Teverovsky, V., & 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.

Tauchen, G. E., & Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica*, *51*(2), 485-505.

Taylor, S. (1986). *Modelling Financial Time Series*. New York: John Wiley & Sons.

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

Teverovsky, V., & Taqqu, M. S. (1997). Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. *Journal of Time Series Analysis*, *18*(3), 279-304.

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

Tsay, R. S. (2001). *Analysis of Financial Time Series*. New York: John Wiley & Sons.

Venables, W. N., & Ripley, B. (2002). *Modern Applied Statistics with S* (4th ed.). Springer-Verlag.

Vetterli, M., & Kovacevic, J. (1995). *Wavelets and Subband Coding*. Englewood Cliffs, N.J: Prentice Hall PTR.

Vuorenmaa, T. A. (2005). *A wavelet analysis of scaling laws and long-memory in stock market volatility* (Research Discussion Papers No. 27/2005). Bank of Finland. Retrieved from `http://ideas.repec.org/p/hhs/bofrdp/2005_027.html`

Whitcher, B. (2012). waveslim: Basic wavelet routines for one-, two- and three-dimensional signal processing [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=waveslim` (R package version 1.7.1)

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

Willinger, W., Taqqu, M. S., & Teverovsky, V. (1999). Stock market prices and long-range dependence. *Finance and Stochastics*, *3*(1), 1-13.

Wuertz, D. (2013). fArma: ARMA Time Series Modelling [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=fArma` (R package version 3.0.1)

Xu, S.-j., & Jin, X.-j. (2009). Predicting drastic drop in Chinese stock market with local Hurst exponent. In *2009 international conference on management science & engineering (16th)* (p. 1309-1315). Moscow, Russia. Retrieved from `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5318022`

Zumbach, G., & Lynch, P. (2001). Heterogeneous volatility cascade in financial markets. *Physica A: Statistical Mechanics and its Applications*, *298*(3), 521-529.

Appendix A

Wavelet methodology

A.1 Frequency domain analysis

A.1.1 Preamble

Most recent studies of financial data adopt the approach from a "time domain" perspective, that is, the data are analysed as time series which are commonly recorded at a pre-determined frequency(s) (i.e. daily, weekly, monthly etc.). This approach, no matter how effective, implicitly limits the recorded frequency as the sole frequency to be considered when studying realizations of a time varying variable. Problems emerge when this assumption turns out to be insufficient. Specifically, what will the situation be when there are many, not one, frequencies that dictate the underlying generating process of the variable of interest?

To address this concern, a new approach taking into account the frequency aspect is proposed. A well-established methodology representing this branch of "frequency-domain" analysis is the Fourier transform/Spectral analysis. In general this method is a very powerful tool specifically designed to study cyclical behaviour of stationary variables, such as those frequently observed in financial data. Based on this fundamental idea, an advanced technique was developed to simultaneously incorporate both aspects-time and frequency-of a data sequence. This relatively novel methodology is known as the wavelet transform. It is worth noting that though wavelet analysis has been well-established in the field of engineering, in particular signal processing, its application in finance is only recently becoming more popular thanks to the effort of pioneers such as R. Gençay, F. Selçuk (deceased) and B. Whitcher.

In the current section we shall present a brief revision of the spectral theory with focus on the intuition and most direct application rather than technical discussion. The technical details are simplified to emphasise only the most essential principles of the theory. With that said various references are provided for readers whose interest lies in more detailed mathematical treatments. The first part of this section serves as the platform for the development of wavelet-based anal-

ysis that inherits the advantages of both time and frequency domain approaches. Later on we shall study in greater detail the construction of wavelet theory and its relevance to our research.

A.1.2 Spectral analysis

A.1.2.1 Detecting seasonality in periodic data

An early motivation for spectral analysis stems from the desire to study economic variables exhibiting cyclical behaviour (for example GDP or inflation rate) by examining the most prominent frequencies at which they fluctuate. It is also crucial that such time series be freed of any seasonalities, as the seasonal components might mask the underlying trend that is of interest to researchers. The cornerstone of Spectral analysis is the spectral density function, or the power spectrum, which shows the strength of the signal's variation (or as engineers would say, the energy) corresponding to a particular frequency. This technique provides a concise inference of the periodic behaviour as well as the cycle length of the time series. To this end, the series of interest, or the "original" signal, (denoted as $\{x(t)\}$) is convolved with a complex exponential operator and is transformed into a function of frequency via the Fourier transform :

$$X(f) = \sum_{t=-\infty}^{\infty} x_t e^{-i2\pi ft} \tag{A.1.1}$$

where f represents the various relevant frequencies, t denotes the time index of the signal, i is the imaginary measure that satisfies $i^2 = -1$. The complex exponential component can be rewritten as in Euler's theorem [43]:

$$e^{-i2\pi ft} = \cos(2\pi ft) - i\,\sin(2\pi ft) \tag{A.1.2}$$

Gençay et al. (2002) pointed out that any infinite but stationary signal can be viewed as a "*combination of an infinite number of sinusoids with different amplitudes and phases*" (p.29). Such a reconstruction of the signal can also be expressed in the form of the inverse Fourier transform:

$$x(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(f)e^{i2\pi ft}df \tag{A.1.3}$$

The relationship between the signal $\{x(t)\}$ and its Fourier transform $X(f)$ (the two of which compose what is known as the 'Fourier pair') is explained via Par-

[43] For a detailed account on complex exponential, see Chiang (1984).

seval's theorem, which essentially claims that the total energy of the signal is obtained by integrating the energy per unit frequency over an interval of 2π (Gençay et al. (2002)):

$$\sum_{t=-\infty}^{\infty} |x_t|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(f)|^2 df \tag{A.1.4}$$

To illustrate this point of view, it is worthwhile to first review the representation of a sinusoid signal in the time domain. This most basic specification is a periodic signal which has the general form of $x_t = A \sin\left(\frac{2\pi t}{p} + \phi\right)$ where $t = 0, 1, 2, \ldots, N-1$. It is a 'stylized' time series with N time points. Here A is its amplitude (or the level at which it fluctuates around zero) which is generally set to 1. The cosine function is related to the sine function via the expression: $\cos(\theta) = \sin(\theta + \pi/2)$. Aside from the previously noted amplitude A and frequency f, we have the phase parameter ϕ (also known as the shift parameter). The cosine curve is said to be a shifted sine curve with the shift parameter of $\pi/2$.

The sinusoid signal is a function of both t and the parameter "period" (p), which denotes the number of time units during which the signal completes a cycle. For example, a signal that repeats the same pattern every 12 months is said to have a period of 12. Of interest is the concept of frequency, which is the inverse of period, and denotes the number of cycles completed per unit time: $f = \frac{1}{p}$. In the context of discrete sampling, we need at least 2 time points to define a cycle ($p \geq 2$). Thus the highest possible frequency for discrete data is $\frac{1}{2}$ which is known as the "Nyquist frequency" or otherwise as "folding frequency" in some documents.

Motivated by the need to identify the prominent periodic components driving the evolution of similar signals, Hamilton (1994) proposed the use of the "population spectrum" of a covariance stationary process. The spectrum is defined as the discrete Fourier transform of the signal's autocovariance [44]:

$$S(\omega) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \gamma_l \, e^{-i\omega l} \tag{A.1.5}$$

in which $\omega = 2\pi f$ is commonly referred to as the "angular frequency". While f is measured by the number of cycles per time unit [45] ω is the number of radians

[44] Equivalently we can write: $S(f) = \sum_{l=-\infty}^{\infty} \gamma_l \, e^{-i2\pi f l}$ where $-\frac{1}{2} \leq f \leq \frac{1}{2}$.

[45] The international measurement unit of frequency is called Hertz (Hz). 1 Hz = 1 cycle per second.

per second. γ represents the autocovariance series in which l signifies the lag parameter. Conversely the autocovariance can be obtained from an inverse Fourier transform of the spectrum.

Cavalcante and Assaf (2004) defined a long-memory process via the behaviour of its spectrum: $\lim_{\lambda_j \to 0^+} S(\lambda_j) = C\lambda_j^{-2d}$. Here the spectrum $S(\lambda_j)$ is estimated at the harmonic frequencies $\lambda_j = \dfrac{2\pi j}{n}$ with $j = 1, 2, \ldots, \dfrac{n}{2}$. C is a positive constant while n is the sample size. This expression effectively captures the slow rate of decay of the autocorrelation function. The link between autocorrelation and spectrum is crucial for our study of long-memory process examined in later chapters. More importantly, this relationship implies whenever the sinusoid associated with a frequency f is highly correlated with γ_l we shall observed large coefficients of the spectrum (Gençay et al. (2002)). Extensive examination of this methodology and its alternatives is beyond the scope of this paper; we are however interested in its basic application in revealing the prominent frequencies of a signal, especially the ones with periodic components.

A.1.2.2 Generalization of Fourier transform

Before going further into the introduction of the periodogram, it would be helpful to generalize the idea of fitting a linear combination of sines (or cosines) at different frequencies to any time series with strong cyclical behaviour (and in some cases, stationarity). Consider the general form of a cosine function representation of a periodic process: $A\ \cos(2\pi ft + \phi)$. A more convenient expression of this function is obtain via the trigonometric identity:

$$A\ \cos(2\pi ft + \phi) = A_1\ \cos(2\pi ft) + A_2\ \sin(2\pi ft) \tag{A.1.6}$$

in which $A = \sqrt{A_1^2 + A_2^2}$, $A_1 = A\cos(\phi)$, $A_2 = -A\sin(\phi)$ and $\phi = \text{atan}\,(-A_2/A_1)$. We can calibrate a regression model based on this expression:

$$x_t = A_0 + A_1\ \cos(2\pi ft) + A_2\ \sin(2\pi ft) \tag{A.1.7}$$

Not surprisingly, such a representation could be interpreted as the simplest form of a trend model with two regressors and is capable of capturing periodic behaviour (if any) of the signal. We can then fit a cosine curve to such signal by estimating a model specified by equation A.1.7, with the cosine and sine terms as regressors and the amplitudes as regression coefficients. As an illustration, with all coefficients significant at the 1% level, Cryer and Chan (2008) pointed out how well the cosine fits a time series of average monthly temperature recorded in Iowa, United States. In addition, if the number of cosine terms in this model is

increased (i.e. adding new regression parameters to it) we shall be able to obtain an even closer fit.

To reduce the intense computation (due to redundancy) of running this regression for large sample size, a modified algorithm known as the Fast Fourier transform has been developed by Cooley and Tukey (1965) (See e.g. Rader and Brenner (1976) for more details). This quick, efficient algorithm is currently incorporated in a wide array of programs, such as the R packages (R Core Team (2013)), allowing us to perform our analysis with relative ease. Next we shall show how to implement this principle with the use of the periodogram.

A.1.2.3 The sample representation of spectrum: the periodogram

With odd sample size N, the periodogram at frequency $\frac{j}{N}$ is defined as the averaged sum of squared amplitudes of the component cosine-sine pairs.

$$P\left(\frac{j}{N}\right) = \frac{N}{2}\left[\hat{A}_{1j}^2\left(\frac{j}{N}\right) + \hat{A}_{2j}^2\left(\frac{j}{N}\right)\right] = \frac{N}{2}\hat{A}_j^2\left(\frac{j}{N}\right) \tag{A.1.8}$$

As we can see from its formation, the periodogram is proportional to the estimate of the original signal's amplitude. Consequently, as Cryer and Chan (2008) put it, "*the height of the periodogram shows the relative strength of cosine-sine pairs at various frequencies in the overall behaviour of the series.*" (p.322). From a different angle, consistent with the derivation of the classical least square estimates, $\hat{A}_{1j}\left(\frac{j}{N}\right)$ and $\hat{A}_{2j}\left(\frac{j}{N}\right)$ essentially represent the correlations of the signal with the sinusoids at frequency $\frac{j}{N}$, so that the periodogram can be thought of as a indicator of how closely each sinusoid, at each frequency, resembles the aggregated signal $\{x_t\}$. When the sample size is even, the periodogram definition is still valid for $j = 1, 2, \ldots, k-1$. However when $j = k$ (or at the Nyquist frequency $f = \frac{k}{N} = \frac{1}{2}$) we have $P(f) = N(\hat{A}_k^2)$.

A.2 Wavelet analysis

Previously we have showed the power of spectral representation based on a localized frequency basis function in modelling stationary and periodic processes. Although the Fourier transform is useful when analysing this type of time series, it would clearly be preferable to allow the amplitudes, or the underlying variances, of the component sinusoids to vary over time, i.e. the presence of non-stationarity. This is accounted for with the wavelet theory. But there is no free lunch: we have to trade the ability to analyse non-stationary time series with the ability to specify

individual frequencies since wavelets work with "bands" of frequencies.

A.2.1 Introduction to wavelets

Wavelet-based methodology can be said to have originated from Haar (1910), although developments in modern wavelet theory date back to the 1980s. To illustrate the approach, we directly describe two interrelated procedures: (i) wavelet decomposition and (ii) its inverse process, wavelet reconstruction, the two of which are crucial for analysing our time series. As such, this section aims to provide a brief and simple introduction to wavelet theory and its most direct applications for our empirical research. Therefore, most of the discussion is very general and more suited to readers whose interest does not lie in the technical area. Others are highly recommended to find more detailed accounts of the methodology from the theoretical works of Daubechies (1992), Percival and Walden (2000) and Mallat (2009), among other authors cited throughout this section.

A.2.1.1 The continuous wavelet functions

The "wavelet" at its core is simply a function, which creates a representation of the original data series by convolving with it. Academics often describe this method as a way to 'project' the data on a function that oscillates on a short time interval. This function, termed the "mother" wavelet, satisfies two basic conditions, as noted by Baqaee (2010):

$$\begin{aligned} \int_{-\infty}^{\infty} |\psi(t)|^2 dt &= 1 \\ \int_{-\infty}^{\infty} \psi(t) dt &= 0 \end{aligned} \tag{A.2.1}$$

The first condition implies that the energy of the function, expressed by a sum of squares, equals one (we say it has unit energy). The second shows that the sum of oscillations is zero. Taken together, we have a "small wave" whose (non-zero) fluctuations die out (or cancel out) quickly. This is in contrast to the infinite persistence of the sine/cosine functions which forms the basis of the Fourier transform (Masset (2008)). In addition, these two conditions are complemented by a fundamental rule that all wavelet functions $\psi(t)$ must satisfy, known as the "admissibility rule" (Goupillaud, Grossmann, and Morlet (1984)):

$$C_\psi = \int_0^{\infty} \frac{|\Psi(f)|}{f} df < \infty$$

where $\Psi(f)$ is the Fourier transform of $\psi(t)$.

To begin our analysis, the mother wavelet must be transformed into a scaled

(dilated) version and then shifted (translated) to a recursive form:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)$$

where s and u are the dilated and translated parameter, respectively. The Continuous wavelet transform (CWT) can now be derived as the projection of function $f(t)$ on the wavelet $\psi_{u,s}(t)$, that is:

$$W(u,s) = \int_{-\infty}^{\infty} f(t)\psi_{u,s}(t)dt$$

By continuously applying this operator to a infinite range of u and s we are able to break the original function down to its simpler components, a process referred to as the "decomposition". It is also possible to reconstruct the function by an inverse operation (Gençay et al. (2002)). Because $W(u,s)$ is redundant, it is impractical and computationally impossible to decompose a discrete signal with all wavelet coefficients from the CWT.

Since our interest mainly lies in the decomposition of a finite time series, we adopt the discrete approach (called the Discrete wavelet transform) in the form of the pyramid algorithm developed by (Mallat (2009)). Gençay et al. (2002) refers to the DWT as some sort of critical sampling of the CWT. This means the DWT contains the minimum number of CWT coefficients sufficient to preserve information of the original signal in a parsimonious way. In the DWT, we have the translated and dilated wavelet function in discrete form:

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$$

where j and k are scale and location parameters, respectively. These parameters determine the length and location of the wavelet over time. One of Daubechies (1992)'s propositions in her book is that a family of all $\psi_{j,k}$ forms an orthonormal basis for the space of all square integrable functions (commonly denoted as $L_2(\mathbb{R})$). This implication is the cornerstone of wavelet theory which allows us to express any function in this space as a "linear combination of multiple wavelets" in a similar manner as decomposing a time series into sinusoids.

The mother wavelet is the key component to our analysis; however, to completely decompose any function f in the space $L_2(\mathbb{R})$ we need a complementary component called the "father" wavelet function (denoted as ϕ). According to Baqaee (2010), with this addition we can represent a function f as follows:

$$f(t) = \sum_{l\in\mathbb{Z}} < f, \phi_l > \phi_l(t) + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} < f, \psi_{j,k} > \psi_{j,k}(t) \tag{A.2.2}$$

where $< ., . >$ denotes the convolution or inner product between the signal and the filters.

The components of the output of DWT is called the wavelet coefficients

$$d(j,k) = 2^{j/2} \sum_t x_t \psi(2^j t - k)$$

and the scaling coefficients

$$s_{j,k} = 2^{j/2} \sum_t x_t \phi(2^j t - k)$$

in which ψ and ϕ are discrete functions. These definitions imply that the wavelet is localized in time and frequency, and the shape of the wavelet coefficient time series will resemble that of the original series. As Kim and In (2010) pointed out, "*coefficients over rough sections or over jumps [...] will be large relative to [...] smooth sections.*" (p.3). In particular, we have larger wavelet coefficients whenever the wavelet function resembles the signal more closely [46]. As a result, by examining these coefficients at individual scales we can identify underlying local fluctuations in volatility series. To sum up, the discrete wavelet transform provides us a tool to look at our data at different horizons, and hereafter we shall refer to this framework as Multi-scale analysis. It allows us to decompose the time series of interest into multiple time horizons, and to look more deeply into the underlying mechanism determining volatility.

A.2.1.2 Multi-scale analysis

In this subsection we provide an introduction to the procedure of a Multi-scale analysis. To begin with, we illustrate the concept with a simple example of a discrete sequence of data observations in the form of a stylized time series:

$$X = (x_1, x_2, \ldots, x_n) \equiv \{X_t, 1 < t < n\}$$

with integer t. Here n is a dyadic real number (which means $n = 2^J$, for a positive integer J). This assumption will be relaxed later on. As noted earlier, the essence of wavelet analysis is to extract information localized both at different scales and different times. These information can be regarded as the 'details' about the "degree of difference or variation of the observations" (Nason (2008)). At the finest scale, this quantity implied information about a particular observation and

[46] In a similar manner, when a sine wave at a particular frequency resembles the signal, the periodogram at that frequency spikes up (see subsection A.1.2.3).

its immediate neighbour:

$$d_k = x_{2k} - x_{2k-1} \quad \text{for} \quad k = 1, 2, \ldots, n/2$$

We can see that unlike the first difference vector, $\{d_k\}$ contains only the difference between non-overlapping adjacent pairs of data points, so that the next detail coefficient after d_k is $d_{k+1} = x_{2k+2} - x_{2k+1}$ and as such, $(x_{2k+1} - x_{2k})$ is missing. For the same reason, $\{d_k\}$ only has half the number of data points compared to the original series. Roughly speaking, $\{d_k\}$ provides us with additional information about the locations $2k - 1/2$. By the construction of this sequence, there is no more information available at this scale. Next we need to construct a new, coarser time series to extract details from. Denoting this series as $\{c_k\}$ we have:

$$c_k = x_{2k} + x_{2k-1} \quad \text{for} \quad k = 1, 2, \ldots, n/2$$

The new sequence is called "scaled local averages" because it is equivalent to a local average series multiplied by 2 (Nason (2008)). Obviously, the new series is a coarser representation of the original, as it contains averages over non-overlapping pairs. Because it is similar to an smoothing moving average operation (except for not dividing the sum), $\{c_k\}$ is also referred to as the "smooths" series [47]. Repeating the process above with $\{c_k\}$ in place of an original series, we are able to derive a new details series as well as construct a smooths series at a coarser scale.

Before moving on, we need some formal notations that specify the scale parameter in our coefficients. Denote the details and smooths series as $\{d_{j,k}\}$ and $\{c_{j,k}\}$ where $1 < j < J$ with $J = \log_2 n$. This helps reminding us that the coefficients at a particular scale are computed from a series which has half the number of observations as the immediate finer scale (a process called 'down sampling'). Roughly speaking, we can see that the fundamentals of "multi-scale transform" here is nothing more than taking "difference of difference" and "average of average" to move from finest to coarsest representations (in the form of details) of the original signal, while preserving informations localized in time (in a similar manner as looking at things through a magnifying glass). The finest to coarsest details/smooths are thus:

$$\begin{cases} \{d_{1,k}\}, \{d_{2,k}\}, \ldots, \{d_{J,k}\} \\ \{c_{1,k}\}, \{c_{2,k}\}, \ldots, \{c_{J,k}\} \end{cases} \tag{A.2.3}$$

[47] Throughout the course of this paper we refer to the following pairs of terms interchangeably: "details/smooths", "detail/smooth coefficients", "detail/smooth series (vectors)".

The above algorithm ends at the coarsest scale, when there is only a single coefficient produced: $c_{J,1}$. To illustrate, we use this algorithm to decompose a discrete time series with 16 observations $\{X_t\}$ as follows:

Level (j)	X_t	3	3	9	11	4	10	10	8	14	19	5	2	1	7	7	3
1	$d_{1,k}$	0		2		6		-2		5		-3		6		-4	
	$c_{1,k}$	6		20		14		18		33		7		8		10	
2	$d_{2,k}$		14				4				-26				2		
	$c_{2,k}$		26				32				40				18		
3	$d_{3,k}$				6								-22				
	$c_{3,k}$				58								58				
4	$d_{4,k}$								0								
	$c_{4,k}$								116								

Table A.1 Simple example of multi-scale decomposition algorithm.

As it turns out, the detail/smooth coefficient vectors described above also go by other terminologies among academics. Again, for the sake of clarification and connection to wavelet functions, in this paper we shall refer to them as "wavelet/scaling coefficients" which is more standard in the literature.

Now we come back to our story with wavelet theory. The difference and average operators we apply are representations of the mother and father wavelet functions in a discrete context. As shown in the next section, this algorithm is a simple case of the Discrete Wavelet transform (DWT). This "pyramid" algorithm (referring to the fact that the operators reduced the number of observations in each iteration) was first generalized by Mallat (2009).

Up to this point we have derived the basics of the one sided multi-scale transformation. From its result we can work backwards and perform the other part of the transformation, i.e. reconstructing the original series using the inverse process. This starts from the coarsest coefficient set ($c_{4,1}$ and $d_{4,1}$). Note that this process only uses the scaling and coarsest wavelet coefficients as inputs. In other words, we reconstruct the signal from its local averages:

$$c_{j,2k} = \frac{c_{j+1,k} + d_{j+1,k}}{2}$$

$$c_{j,2k-1} = \frac{c_{j+1,k} - d_{j+1,k}}{2}$$

Applying this process to the output of the previous example returns the original series, as follows:

Level (j)																	
4	$c_{4,k}$	116															
	$d_{4,k}$	0															
3	$c_{3,k}$	58								58							
	$d_{3,k}$	6								-22							
2	$c_{2,k}$	26				32				40				18			
	$d_{2,k}$	14				4				-26				2			
1	$c_{1,k}$	6		20		14		18		33		7		8		10	
	$d_{1,k}$	0		2		6		-2		5		-3		6		-4	
	X_t	3	3	9	11	4	10	10	8	14	19	5	2	1	7	7	3

Table A.2 Simple example of multi-scale reconstruction algorithm.

Remark There exists interchangeable usage of the terms *level* and *scale* in the literature and it is a source of confusion. We try to avoid mixing up the two by assigning 'level' with j and 'scale' with 2^j or scale length $\lambda_j = 2^{j-1}$. Specifically, scale parameter λ_j implies that the details/smooths coefficient vectors at level j have length $n/2^j$ and are computed from a series with length $n/2^{j-1}$. Thus the highest level denotes the coarsest scale. Note that there is an inversion of the order of levels when comparing the decomposition and reconstruction process, as illustrated in the previous example.

A.2.2 Discrete wavelets transform

A.2.2.1 Haar wavelets

This section is dedicated to generalizing the nature of the filters we used in the previously described algorithm. Continuing the example in subsection A.2.1.2, we see that the output of the DWT is 3 sequences of wavelet coefficients and only 1 scaling coefficient at the highest level (as all other scaling coefficient vectors are used as input of the decomposition at the next level). To be specific, with a 16-elements input vector

$$I = \{X_t, t \in (1, 16)\} = (3, 3, 9, 11, 4, 10, 10, 8, 14, 19, 5, 2, 1, 7, 7, 3)$$

we have an equal length output vector:

$$O = \{d_j, c_J | 1 \leq j \leq 4\} = (0, 2, 6, -2, 5, -3, 6, -4, 14, 4, -26, 2, 6, -22, 0, 116)$$

An important note is that this transformation does not preserve the signal's energy (expressed as sum of squared values) since:

$$||X_t^2|| = \sum_{i=1}^{16} x_i^2 \neq \left(c_J^2 + \sum_{j=1}^{3} d_j^2\right)$$

To compensate for this we need to modify the differencing and averaging operators as follows:

$$d_k = \frac{x_{2k} - x_{2k-1}}{\sqrt{2}}; \quad c_k = \frac{x_{2k} + x_{2k-1}}{\sqrt{2}}$$

so that original signal's energy is now preserved at each iteration of the transformation (or the output of the transformation inherits the same overall variance as the signal):

$$d_k^2 + c_k^2 = \left(\frac{x_{2k} - x_{2k-1}}{\sqrt{2}}\right)^2 + \left(\frac{x_{2k} + x_{2k-1}}{\sqrt{2}}\right)^2 = x_{2k}^2 + x_{2k+1}^2$$

this could be understand as the result for the first level. Nevertheless it is true at every level. In general, for a J levels decomposition we have:

$$d_1^2 + d_2^2 + \cdots + d_J^2 + c_J^2 = ||X_t^2||$$

To put it the 'formal' way, the modified wavelet filter has a coefficient vector $h = (1/\sqrt{2}, -1/\sqrt{2})$ while its scaling counterpart has a coefficient vector $g = (1/\sqrt{2}, -1/\sqrt{2})$. These length $L = 2$ pairs of filters constitutes the most basic form of the Haar filters [48]. To be more specific, the Haar discrete filters can be expressed by their functional form as:

$$\psi \sim h_l = \begin{cases} 1/\sqrt{2} \text{ for } l = 0 \\ -1/\sqrt{2} \text{ for } l = 1 \\ 0 \text{ otherwise} \end{cases} \qquad \phi \sim g_l = \begin{cases} 1/\sqrt{2} \text{ for } l = 0 \\ 1/\sqrt{2} \text{ for } l = 1 \\ 0 \text{ otherwise} \end{cases} \tag{A.2.4}$$

Gençay et al. (2010) pointed out that they possess three important properties, analogous to the conditions the mother and father wavelet functions satisfy in (A.2.1). In general, for a length L pair of filters:

- $\sum_{l=0}^{L-1} h_l = 0$ and $\sum_{l=0}^{L-1} g_l = \sqrt{2}$: this makes the function oscillates.

[48] Named after the Hungarian mathematician Alfred Haar (Haar (1910)).

- $\sum_{l=0}^{L-1} h_1^2 = 1$ and $\sum_{l=0}^{L-1} g_1^2 = 1$: this is called unit energy, which ensures the oscillation is short-lived. It also allows for energy preservation.
- $\sum_{l=1}^{L} h_l h_{l+2n} = 0$ and $\sum_{l=1}^{L} g_l \; g_{l+2n} = 0 \;\; \forall \; n \neq 0$: this means the filters are orthogonal to their even shifts.

In addition, for each level, the elements of the wavelet and scaling filters satisfy the quadrature mirror relationship $g_l = (-1)^{l+1} h_{L-1-l}$ for $l = 0, \ldots, L-1$ (See Vetterli and Kovacevic (1995). Applying the Haar DWT to the previous example yields:

Level (j)	X_t	3	3	9	11	4	10	10	8	14	19	5	2	1	7	7	3
1	$d_{1,k}$	0		1.414		4.242		-1.414		3.535		-2.121		4.242		-2.828	
2	$d_{2,k}$			7				2				-13				1	
3	$d_{3,k}$					2.121								-7.778			
4	$d_{4,k}$							-5.603 $\times 10^{-16}$									
	$c_{4,k}$							29									

Table A.3 Simple example of multi-scale decomposition with Haar DWT.

A.2.2.2 Generalization with matrix representations

The role of the orthonormal basis From the previous discussion we know wavelet transform algorithm has three components: the original signal (input), the coefficients vector (output) and a filter defining the DWT. The filter extracts all information (at different time points) from the signal and expresses it as coefficients associated with high/low frequencies [49]. In our example, a collection of orthonormal basis functions (which is a composition of the wavelet filter h_l and scaling filter g_l) forms a $n \times n$ orthogonal matrix $\mathcal{W}$:

$$\mathcal{W} = \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{W}_2 \\ \vdots \\ \mathcal{W}_J \\ \mathcal{V}_J \end{bmatrix} \tag{A.2.5}$$

in which $\mathcal{W}_j$ and $\mathcal{V}_J$ are the sub-matrices constructed based on level j wavelet filter and level J scaling filter, respectively. For example, consider $\mathcal{W}_1$:

[49] Unlike the infinite length sinusoids used in Fourier transform, wavelet's basis functions are localized in time and thus the coefficients produced are allowed to be time-varying.

$$\mathcal{W}_1 = \left[h_1^{(2)}, h_1^{(4)}, \ldots, h_1^{(n/2-1)}, h_1\right]^T$$

where h_1 is a length n zero-padded wavelet filter coefficient vector at level 1. $h_1^{(2)}$, $h_1^{(4)}$ ... $h_1^{(n/2-1)}$ are acquired by shifting h_1 circularly by a factor of 2, 4 ... $n/2-1$:

$$\begin{aligned} h_1 &= [0,0,0,0,\ldots,0,0,h_{1,1},h_{1,0}]^T \\ h_1^{(2)} &= [h_{1,1},h_{1,0},0,0,0,\ldots,0,0]^T \\ h_1^{(4)} &= [0,0,h_{1,1},h_{1,0},0,\ldots,0,0]^T \\ &\vdots \\ h_1^{(n/2-1)} &= [0,0,0,\ldots,h_{1,1},h_{1,0},0,0]^T \end{aligned} \tag{A.2.6}$$

repeat the process with other levels we obtain $\mathcal{W}_2, \mathcal{W}_3, \ldots, \mathcal{W}_J$ and $\mathcal{V}_J$.

From the previous example we see that a dyadic finite length n signal $\{X_t, t \in (1, \ldots, n = 2^J)\}$, using the Haar DWT, we can produce the corresponding length n coefficient vector denoted as $O = (\{d_{j,k}\}, c_{J,1})$. This vector can also be rewritten in a more compact form:

$$\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_J, \mathbf{V}_J)^{\mathrm{T}} \tag{A.2.7}$$

where $\mathbf{W}_j$ is the wavelet coefficient vector at level j and $\mathbf{V}_J$ is the scaling coefficient vector at level $j = J$ [50]. We have $\mathbf{W} = \mathcal{W}X_t$. Since $\mathcal{W}$ is orthogonal it follows that:

$$||\mathbf{W}^2|| = \mathbf{W}^T\mathbf{W} = (\mathcal{W}X_t)^T\mathcal{W}X_t = X_t^T(\mathcal{W}^T\mathcal{W})X_t = X_t^T X_t = ||X_t^2||$$

so that the Haar DWT preserves the signal's energy and length (it is not true, however, that all DWT are based on orthogonal transform). This special feature allows us to perform an analysis of variance in the frequency domain, and effectively study each level's contribution to the overall variance of the process via the following relationship presented in Gençay et al. (2002):

$$||X_t^2|| = \sum_{j=1}^{J} ||\mathbf{W}_j||^2 + ||\mathbf{V}_J||^2$$

where $||\mathbf{W}_j||$ denotes the portion of energy of the signal as a result of variations at scale λ_j, this quantity is proportional to the total variance of the process.

[50] Here we suppress the location/time parameter. The formal notations should be $\mathbf{W}_{j,k}$ and $\mathbf{V}_{J,k}$ with $k = 1, 2, \ldots, n/2^j$.

A.2.2.3 Multi-resolution analysis (MRA) with the DWT

Deriving the wavelet/scaling vector a crucial step in a wavelet-based analysis. The vector provides information related to local differences and averages of the original signal. Nevertheless, our research requires a representation of the signal at each level. To do this, we need to reconstruct the series at each level, using the wavelet coefficients and the orthogonal basis at that scale. The process is described as follows:

- After decomposing the signal with DWT we obtain the vector $\mathbf{W}$.
- For level j (scale λ_j) compute the details coefficient vector as: $D_j = \mathcal{W}_j^T \mathbf{W}_j$ where $\mathbf{W}_j = \mathcal{W}_j X_t$ is the level j wavelet coefficients.

Here D_j can be thought of as a 'rough' approximation of the signal at level j. With a total of J wavelet vectors we can derive J time series of details. Similarly, we have $S_J = \mathcal{V}_J^T \mathbf{V}_J$ denotes the "residual" component of the signal after extracting all the details at lower levels. This sequence is called 'smooth' coefficients [51]. Note that unlike $\mathbf{W}_j$ and $\mathbf{V}_J$, D_j and S_J have the same length as X. As Gençay et al. (2010) pointed out, the MRA can be defined as an additive decomposition of the signal:

$$X_t = \sum_{j=1}^{J} D_{j,t} + S_{J,t}$$

Also, we have $S_j = \sum_{k=j+1}^{J} D_k$, meaning the smooth series contains all of the detail (or variations) at higher scales (which are associated with lower frequencies). More notably, as all variations associated with higher frequencies have been extracted, the smooth series at the lowest frequency, S_J, is essentially a sample mean of the signal. Note that at each level, the smooth series S_j is a representations of the signal's underlying trend, analogous to a moving average series. In this regard, the MRA effectively breaks the signal into several components including trend, seasonality and irregulars, a feat performed by various tool of basic time series theory as well as the Fourier methodology.

Before moving on, we need to emphasize another important feature of the orthonormal DWT that is crucial in wavelet variance analysis. As noted in subsection A.2.2.2, the orthogonality of the wavelet basis matrix allows us to decompose the signal's variance using the wavelet and scaling vectors. It can be shown that a similar feat is possible with the detail and smooth coefficient vectors. As $D_j = \mathcal{W}_j^T W_j$ and $S_J = \mathcal{V}_J^T V_J$, we also have $D_j^T D_j = W_j^T W_j$ and $S_J^T S_J = V_J^T V_J$

[51] The terms "detail" and "smooth" were originally adopted by Percival and Walden (2000).

so that:

$$||X_t||^2 = \sum_{j=1}^{J} ||\widetilde{W}_j||^2 + ||\widetilde{V}_J||^2 = \sum_{j=1}^{J} ||D_j||^2 + ||S_J||^2$$

to put it differently, "*[...] we have an alternative representation for the variability of a time series via its squared wavelet coefficients.*" (Gençay et al., 2002, p.236).

Remark Here we summarize various interrelated indicators of the coefficient vectors associated with wavelet analysis, for the purpose of clarification:

- wavelet/scaling filter vectors $h_{j,l}/g_{j,l}$ ($j = 1, \ldots, J$; $l = 0, \ldots, L-1$): these define the filtering operations and are components of the DWT.
- wavelet/scaling coefficient vectors $W_{j,k}/V_{j,k}$ ($j = 1, \ldots, J$; $k = 1, \ldots, n/2^j$): these define outputs of the DWT.
- detail/smooth coefficient vectors $D_{j,t}/S_{j,t}$ ($j = 1, \ldots, J$; $t = 1, \ldots, n$): these define outputs of the decomposition process. Recall that for level j we have $D_j = \mathcal{W}_j^T \mathbf{W}_j$ where $\mathbf{W}_j$ is the wavelet coefficient vector and $\mathcal{W}_j$ is the matrix representing the DWT.

.

A.2.3 Maximal overlap discrete wavelet transform

An important generalization (in a sense) of the DWT is the Maximal overlap discrete transform (MODWT). Gençay et al. (2002) presents a detailed matrix formulation of the MODWT. Similar to the DWT, the length n wavelet coefficient vector $\widetilde{W}$ is computed via:

$$\widetilde{\mathbf{W}} = \widetilde{\mathcal{W}} X_t \quad \text{where} \quad \widetilde{\mathcal{W}} = \begin{bmatrix} \widetilde{\mathcal{W}}_1 \\ \widetilde{\mathcal{W}}_2 \\ \vdots \\ \widetilde{\mathcal{W}}_J \\ \widetilde{\mathcal{V}}_J \end{bmatrix} \tag{A.2.8}$$

Here $\widetilde{\mathcal{W}}_j$ is the $n \times n$ sub-matrix constructed by circularly shifting the rescaled filters $\tilde{h}_j = h_j/2^j$ and $\tilde{g}_J = g_J/2^J$ by a factor of one:

$$\widetilde{\mathcal{W}_j} = \left[\tilde{h}_j^{(1)}, \tilde{h}_j^{(2)}, \ldots, \tilde{h}_j^{(n-1)}, \tilde{h}_j\right]$$

Recall that for the DWT, we construct $\mathcal{W}_j$ by shifting the filters by a factor of two. This means we do not practice down sampling in MODWT filtering process, so that the wavelet and scaling coefficient vectors are redundant (each has length n at every level). After computing $\widetilde{\mathbf{W}} = \widetilde{\mathcal{W}} X_t$ the output coefficient vector is

$$\widetilde{\mathbf{W}} = (\widetilde{\mathbf{W}}_1, \widetilde{\mathbf{W}}_2, \ldots, \widetilde{\mathbf{W}}_J, \widetilde{\mathbf{V}}_J)^T$$

where the column sub-vectors $\widetilde{\mathbf{W}}_j$ and $\widetilde{\mathbf{V}}_J$ all have length n rather than $n/2^j$. Note that the $n \times n$ basis matrix $\widetilde{\mathcal{W}}$ is no longer orthonormal. Thus the output vector $\widetilde{\mathbf{W}}$ is highly redundant as it has $(J+1)n$ rather than n coefficients.

Unlike the DWT, we cannot rely on the detail and smooth coefficients to perform an analysis of variance any more, although we can still use the wavelet and scaling coefficients (Percival and Walden (2000)):

$$||X_t^2|| = \sum_{j=1}^{J} ||\widetilde{W}_j||^2 + ||\widetilde{V}_J||^2 \neq \sum_{j=1}^{J} ||D_j||^2 + ||S_J||^2$$

The energy preservation feature of DWT does not transfer to MODWT because instead of having unit energy, the MODWT filters now satisfy:

$$\sum_{l=0}^{L_j-1} \tilde{h}_{j,l} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} = \frac{1}{2^j}$$

Strengths of the MODWT Primary weaknesses of the DWT include (i) its use being limited to dyadic length signals and (ii) its shift-variant MRA. This is not the case for the MODWT. Although forfeiting orthogonality, MODWT can handle signals with any length. More importantly, MODWT-based MRA is shift invariant, meaning circularly shifting/translating the signal will cause the MODWT wavelet and scaling vectors to change accordingly, causing no loss of generality [52].

Furthermore, MODWT-based MRA exhibits proper alignment across time throughout all levels. Jensen and Whitcher (2000) pointed out that an ordinary temporal aggregate (such as the DWT) to a finite dataset leads to a decrease of number of observations and corresponding loss of information associated with high frequency behaviour. However, MODWT can preserve these information thanks to its redundancy and all the scale-specific details, and smooths are "lined up in time with the events of the original series".

[52] Which is why the MODWT filters go by many names in the literature: e.g. "translation-invariant" (Liang and Parks (1996)), "time-invariant" (Pesquet, Krim, and Carfantan (1996)) or "stationary"(Nason and Silverman (1995)).

The advantages of MODWT are by no means trivial and are crucial for studying features of signals in time-frequency domain. MODWT thereby plays a central role in many empirical studies and is the primary tool in our paper. Thereafter whenever a wavelet-based analysis is adopted, the method used will be MODWT unless explicitly stated otherwise. In later Chapters we show how the time and frequency localization aspect of MODWT is ideal to handle both stationarity and time-varying behaviour of non-stationary processes. Jensen and Whitcher (2000) asserts that this property of MODWT makes it ideal for analysing the intradaily patterns (e.g. time-of-the-day effects and irregular occurrences of market fluctuations) and interdaily long-memory exhibited by financial volatility time series.

A.2.4 The choice of an appropriate wavelet filter

Previous studies indicate that there are several factors that need to be considered when choosing a suitable class of discrete wavelet filter for our analysis. Unfortunately there is no universal guideline for the selection. A researcher is then required to weight the strengths and weaknesses of each filter carefully while keeping in mind the research object to select the most appropriate candidate.

For one thing, length of the filter determines the effectiveness of solution to boundary conditions and vanishing moments issues. In principle, it is preferable to decompose longer time series with a longer filter. Also, the degree to which the filter approximates an ideal band-pass filter is improved with increasing filter length (Gençay et al. (2002)). Aside from the simplest (shortest) and only symmetric compactly supported filter, which is the Haar filter, most researchers choose one of the two from the Daubechies class filters, the Extremal phase (EP) and Least asymmetric (LA). In addition, we need to choose a wavelet whose smoothness and symmetry are suitable to analyzing financial time series. One prefers the Daubechies wavelets over other wavelets, for example the Morlet family, since they are more suitable to deal with non-stationary data, as discussed earlier. Finally, we also need to note that the longer the filter is, the more serious the boundary conditions and the more complicated our computations will be. An excellent discussion on the matter as well as review of multiple wavelet-based applications in economics and finance can be found in Crowley (2007).

In general, an ideal wavelet basis for studying financial time series should be sufficiently long, symmetric and smooth. For these reasons a major portion of existing literature features the LA(8) MODWT wavelet which handles the boundary conditions using the 'periodic' method. It is therefore our primary choice unless explicitly stated otherwise. The implementation of various wavelet methodolo-

gies described in this Appendix is made primarily via the R package waveslim introduced by Whitcher (2012).

A.2.5 Application of wavelets on filtering seasonality

We conclude this Appendix by illustrating the ability of an LA(8) MODWT in filtering out the seasonality that could distort the interpretation of financial models. This is because such periodic behaviour may mask the underlying trend of the data. As financial data in general and high-frequency data in specific tend to exhibit strong seasonality, a significant effort was dedicated to the search for methods that are capable of filtering out this effect (see e.g. Muller et al. (1990), Dacorogna, Muller, Nagler, Olsen, and Pictet (1993) and Andersen and Bollerslev (1997a)). The debate is highly relevant to our research, as our high-frequency volatility is also an intraday seasonal series. Gençay et al. (2002) proposed a new method utilizing the wavelet-based MRA. This method enjoys the advantage of not depending on specifications of model or parameter. The principle to extract seasonalities can be generalized into three steps: (i) identify the prominent frequencies that governs the periodic components of the data; (ii) perform the MRA up to a level corresponding to the lowest prominent frequency; (iii) the smooth series from this MRA is a good approximation of the data's underlying trend [53].

Consider a simple illustrative signal that is governed by (a) an AR(1) process plus (b) cyclical terms and (c) a random noise (individually, these components each represents an aspect of a typical time series: (a) trend, (b) seasonality(s) and (c) innovations) :

$$x(t) = 0.9\ x_{t-1} + 7\ \sin\left(\frac{2\pi t}{4}\right) + 7\ \sin\left(\frac{2\pi t}{5}\right) + 7\ \cos\left(\frac{2\pi t}{6}\right) + u_t \qquad \text{(A.2.9)}$$

in which u_t is a white noise process that is normally distributed with mean zero and unit variance. To simplify the example we set all the amplitude terms of the sine and cosine waves to 7. A simulated set of $n = 100$ observations of $\{x(t)\}$ is generated and its autocorrelation function computed. The wavelet-based filtering method can be applied to this time series. Results are reported in Figure A.1.

[53] Note that in their example with foreign exchange data, (Gençay et al. (2002), p.146-149) compute the smooth series after performing a level 8 MRA. Details from level 1 to 8 capture every oscillations with period length ranges from 2 to 512 (corresponding to a frequency band of $1/512 < f < 1/2$). Because their signal includes 288 observations per day, the smooth series should be able to exclude all intraday seasonalities, leaving the trend.

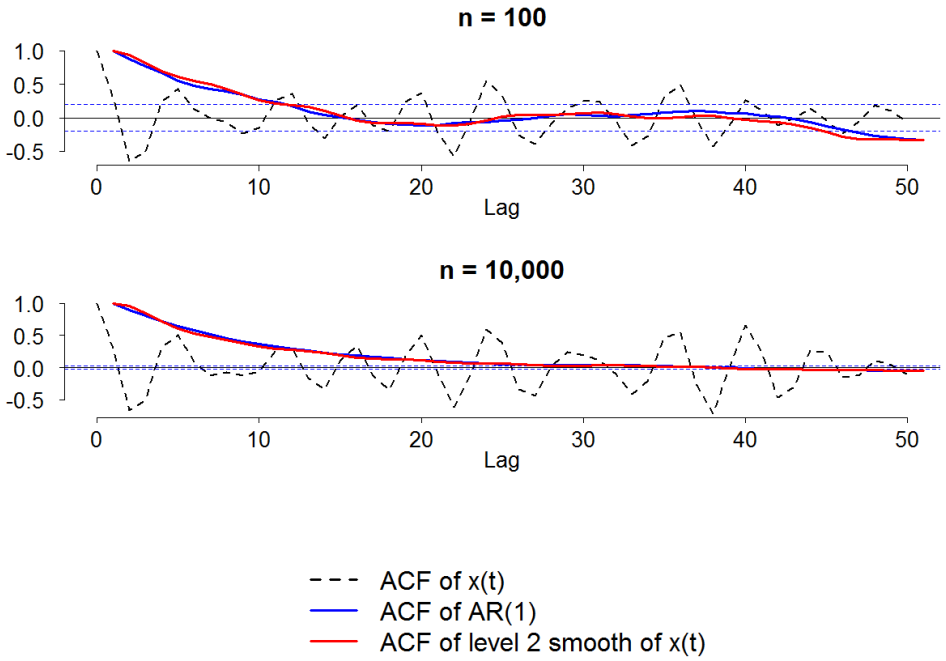


Figure A.1 Autocorrelation functions of (1) the original signal $x(t)$ (dashed line), (2) a sample of AR(1) process (blue line) and (3) a level 2 smooth series (red line). Blue dashed lines indicate the 95% confidence interval.

As can be seen, the ACF of the original signal x_t hardly reveals any sign of the trend component. In particular, while an AR(1) process with persistent parameter $\rho = 0.9$ exhibits a slow decaying ACF, the ACF of the signal oscillates around zero. Now, a level 2 MRA is expected to yield a smooth series capable of capturing fluctuations associated with frequency $f < 1/8$. As the frequencies corresponding to periodic components ranges from $1/4$ to $1/6$, the constructed smooth series should filter out most periodic behaviours and expose the trend. Indeed, when we compare the ACF of the smooth series with that of the underlying trend AR(1) there is little difference. In addition, when we repeat the experiment with an extended signal ($n = 10,000$) the ACF of the signal decays much more slowly, confirming the long-memory nature of the simulated data when seasonalities are filtered out. This simple example shall be extended in Chapter 4, when we study the intraday seasonality exhibited by the high-frequency returns time series.

Appendix B

Alternative long memory estimators

This Appendix describes several techniques related to the R/S method for estimating the intensity of long range dependence in a time series, as suggested in the theoretical review of Taqqu et al. (1995).

B.1 The Aggregated variance method (aggVar for short)

To apply this technique, the time series $\{X_t, t \in (1, 2, \ldots, n)\}$ is divided into m ranges of size k ($m = 1, 2, \ldots, \frac{n}{k}$). We then take an average within each range to create a sample of the aggregated process $\{X_k(m)\}$, with the mean value of the m-th range defined as:

$$X_k(m) = \frac{1}{k} \sum_{t=(m-1)k+1}^{mk} X_t \quad \text{where} \quad m = 1, 2, \ldots, \frac{n}{k} \tag{B.1.1}$$

As shown by Dieker and Mandjes (2003), because $\{X_k\}$ is self-similar, for a large k the distributions of $\{X_k\}$ and $k^{H-1}X_k$ are identical. Thus we have $\mathrm{Var}[X_k] = k^{2H-2}\,\mathrm{Var}(X_k)$. For a given k, the sample variance of $\{X_k\}$ is used as an estimator of $\mathrm{Var}[X_k]$:

$$\widehat{\mathrm{Var}[X_k(m)]} = \frac{1}{n/k} \sum_{m=1}^{n/k} \left(X_k(m) - \overline{X(m)}\right) = \frac{1}{n/k} \sum_{m=1}^{n/k} X_k^2(m) - \left[\frac{1}{n/k} \sum_{m=1}^{n/k} X_k(m)\right]^2 \tag{B.1.2}$$

boIn this way we obtain a different $\widehat{\mathrm{Var}[X_k(m)]}$) for each different range size k. Finally we plot the logarithm of $\widehat{\mathrm{Var}[X_k(m)]}$ against that of the varying k. When the variance estimates are unbiased, the fitting line will have a slope of $\beta = 2H-2$ ($-1 \leq \beta < 0$). The case when $\beta = -1$ is corresponding to a short-range (or zero) dependence.

It is worth noting that in the presence of non-zero correlation because of long-range dependence, the estimate in equation B.1.2 is biased when $\frac{n}{k}$ is small,

or k is large. Specifically:

$$E\Big(\widehat{\mathrm{Var}[X_k(m)]}\Big) = \mathrm{Var}[X_k(m)](1 - \frac{n}{k}\, C^{2H-2}) \tag{B.1.3}$$

for a positive constant C. Therefore when there exists long-memory, or $H > 0.5$, we have $\dfrac{E\Big(\widehat{\mathrm{Var}[X_k(m)]}\Big)}{\mathrm{Var}[X_k(m)]} < 1$ or the sample variance underestimates the true variance, and the resulted estimate of H also suffers from a downward bias. The bias is larger with smaller m (or larger k) but slowly disappears when m is large (see e.g. Beran (1994)). This issue exists with any estimating methods based on a log-log plot such as those described in the current subsection and afterwards. Applying the aggVar method to our simulated fGn process gives us the estimate value of 0.6928. The result log-log plot is shown in Figure B.1. As expected, the Hurst index is underestimated in this case.

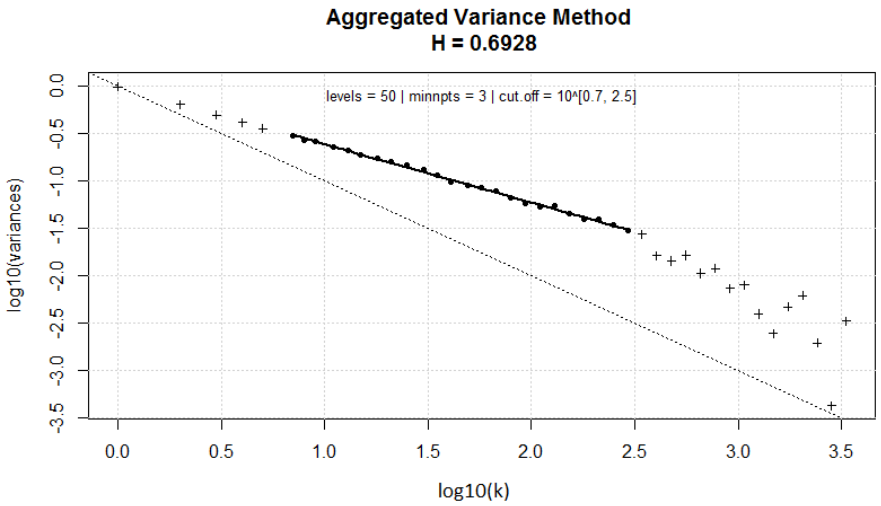


Figure B.1 Log-log plot for estimated Hurst index from an fGn process with $H = 0.7$ by the aggVar method. The dotted line corresponds to the case when $H = 0.5$ which has the slope of -1.

Remark When fitting a line through the log-log points, we need to specify three parameters:

- m: the total number of ranges/blocks or aggregation levels from which the variances or moments are computed.
- Minimum number of observations, or range size, to be used to estimate variance at any aggregation level.
- The fitting range: the range contains the points that are used for fitting the line. This range is specified with the so-called "cut-off" values on a $\log_{10}$ scale. These values should be chosen to define a linear range.

The choice of these parameters is somewhat arbitrary, as there is no explicit method to justify the goodness of H estimates for differing choices. Generally we have to choose a range so that we can fit a relatively linear line through the points. We opt for the default values specified in the R-package fArma (Wuertz (2013)). That is, m is set to 50, minimum range size is 3 and fitting range is from 0.7 to 2.5 (meaning $10^{0.7} < k < 10^{2.5}$). To our knowledge the estimate of H varies little when we change the values of m and k. However, the estimate becomes more and more downward biased as the fitting range increases. As can be seen from Figure B.1, if we include the points on the bottom right, the fitting line will have a lower slope. Also, fitting a line through these points is less accurate. This is the problem with large k (or small m) as described in equation (B.1.3). This also justifies the choice of the fitting range (0.7 to 2.5) as it excludes small and large values of k although with a simulated process like ours, the unbiasedness for small values of k is minimal.

Other variance-based methods To account for the presence of non-stationary behaviours such as jumps in mean and slowly declining trends, Teverovsky and Taqqu (1997) propose the use of two derivatives of the aggVar method:

First, the Differenced variance method (diffaggVar for short) in which they use

$$\widehat{\mathrm{Var}X_{k_{i+1}}} - \widehat{\mathrm{Var}X_{k_i}}$$

in substitution for the sample variance series.

Second, the Absolute Aggregated variance method (AbaggVar for short): the only difference between this method and the aggVar is that instead of calculating sample variance, they use the sum of absolute aggregated series, also called the

absolute moments, which is defined as:

$$AM_k = \frac{1}{n/k} \sum_{m=1}^{n/k} |X_k(m) - \overline{X(m)}|^n$$

We know that for large k, self-similarity implies that $X_k(m)$ and $k^{H-1}X_k(m)$ are identically distributed. For some $n > 0$ and constant C, Dieker and Mandjes (2003) speculate that the following holds:

$$E(AM_k) \sim k^{n(H-1)} E|X_k - E(X_k)|^n \Big(1 - C\ m^{n(H-1)}\Big) \ \text{ as } \ m \to \infty$$

so that $E(AM_k)$ is proportional to $k^{n(H-1)}$. Consequently, the slope of the line obtained by regressing the logarithm of this statistic against that of k shall have a value of $\beta = H - 1$. Note that the method is generally applied with $n = 1$ and for $n = 2$ it reverts to the aggregated variance method. Using the same choice of parameters as in the case of the aggVar, log-log plot of this method is reported in the following figure, offering a comparable estimate of 0.702.

Aside from these variance-based methods, we also adopted two popular alternatives introduced by Taqqu et al. (1995): Higuchi's method and the Residuals of regression method.

B.2 Higuchi's method (Higuchi for short)

Similar to the absolute aggregated variance method above, Higuchi (1988) suggested calculating the partial sums of the original fGn $\{X_t, t = 1, \ldots, n\}$ to produce an fBm:

$$\begin{aligned} Y(t + mk) &= \sum_{t=1}^{t+mk} X_t \\ Y\left[t + (m-1)k\right] &= \sum_{t=1}^{t+(m-1)k} X_t \end{aligned} \tag{B.2.1}$$

then computing the normalized length of the curve:

$$L(m) = \frac{N-1}{m^3} \sum_{t=1}^{k} \frac{1}{m_t} \sum_{m=1}^{m_t} \Big| Y(t + mk) - Y\left[t + (m-1)k\right] \Big| \tag{B.2.2}$$

with m_t is the integer part of $(n - t)/k$. Next we plot $L(m)$ against k on the log-log plot. As $E\left[L(m)\right] \sim Ck^{-h}$, the fitted line will have a slope of $h = 2 - H$. This method gives us an estimated value of 0.6715.

B.3 Residuals of regression method (Peng for short)

This method was originally proposed by Peng et al. (1994), which is commonly referred to as Peng method: Within each range of size k we regress the partial sums on a line given by $\hat{\alpha}_k + \hat{\beta}_k i$. Then we obtain the regression residuals:

$$\epsilon_k(i) = \sum_{t=mk}^{km+i-1} X_t - \hat{\alpha}_k - \hat{\beta}_k i$$

subsequently, the sample variance of the residuals are computed for each range. By averaging this variance across all ranges we get a quantity that is equivalent to the sample variance of the whole series, and is proportional to k^{2H} for the case of fGn and FARIMA processes (this is proved in the Appendix of Taqqu et al. (1995)). Thereby a log-log plot yields a line with slope $2H$. This technique is extensively studied by Cannon, Percival, Caccia, Raymond, and Bassingthwaighte (1997) who refer to it as the "scaled windowed variance" method. Applying it to our series gives an estimate of $H = 0.6956$.

B.4 Periodogram method (Per for short)

Recall from section A.1.2.3 of Chapter A, the periodogram function of a time series $\{X_j\}$ is defined as

$$P(f) = \frac{1}{2\pi n}\Big|\sum_{j=1}^{n} X_j e^{ijf}\Big|$$

with frequency f, length of series n, $i = \sqrt{-1}$. After computing the periodogram $P(f_k)$ $(k = 1, \ldots, n)$ for each frequency $f_k = \dfrac{\pi k}{n}$ we can plot those against frequencies on a log-log scale. The line (similar to the ones we have covered in other methods) should have a slope of $1 - 2H$. Following Taqqu et al. (1995), we also use the lowest 10% of the frequencies to fit the line (hence the negative value of the logarithmic transform in Figure B.2). This method yields an estimate of 0.7057.

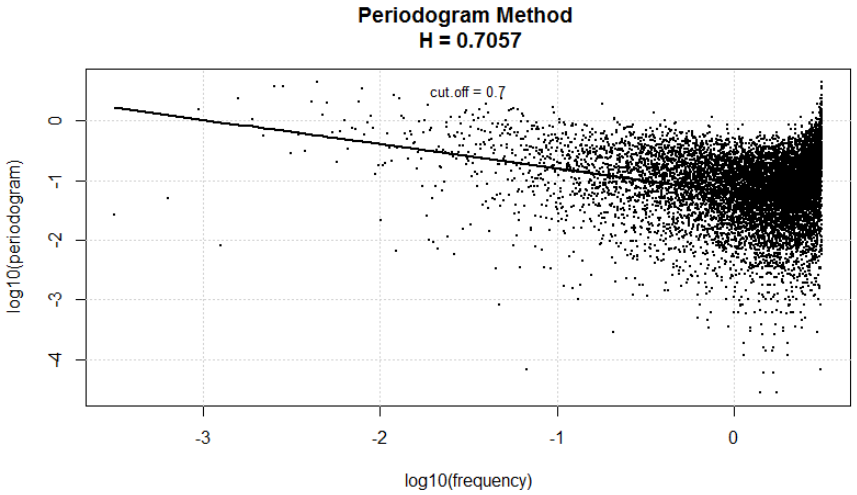


Figure B.2 Log-log plot for estimating Hurst index from a fGn process with $H = 0.7$ by Per method.

As we can see the periodogram decays roughly linearly for this particular simulated data set. Since a periodogram is simply a representation of autocorrelation function on a double logarithmic scale, this result supports the power-law relation characterizing a long-memory process, i.e. $\rho(l) \sim c_\gamma |l|^{-\alpha}$ (see e.g. Beran (1994)). As the lag increases (or at higher frequencies) the strength of autocorrelation decreases and the periodogram fluctuates heavily around the fitted line.

It should be noted that the Periodogram method is one of the most popular long-memory estimating methods to date. However, a note should be made of the reliability of this method: as Vuorenmaa (2005) pointed out, the periodogram is an inconsistent estimator of the spectrum. Consequently, OLS-based estimators such as the one described by Taqqu et al. (1995) and also the well-known GPH estimator (from Geweke and Porter-Hudak (1983)) are inconsistent estimators of the long memory parameters H and d. Furthermore, the frequencies on the far right of the log-log plot corresponds to most of the values and may cause a bias in the slope of the regression line. To compensate for these weaknesses, several researchers have proposed the use of a refined approximation of the spectral density function known as the Modified periodogram (modPer for short). For the modified method, the frequencies are divided into equally spaced boxes. Averaged periodograms are then computed for frequencies in each boxes, excluding the lowest frequencies. Then we fit a line through the resulting points. Compared to the Per method, we do not obtain a significantly different estimate from this method and thus do not report its results. However, we still use this method when examining the long run dependence of our simulated data and compare its performance with other methods at the end of Chapter 2.

Appendix C

Some models in the GARCH family

C.1 Integrated GARCH

According to Reider (2009), evidence for persistent volatility behaviour can be observed in stock market movements around the crash of October 19, 1987. Before that day the standard deviation of equity returns was approximately 1%, then on that day the S&P500 index dropped an astounding 20%. Afterwards volatility increased a little before eventually adjusting back to the pre-crisis level (See Figure C.1). How quickly volatility stabilizes can be determined by estimating the persistence parameter from the GARCH model. Most empirical studies have recorded this parameters in close proximity to 1, supporting the hypothesis that the ARMA polynomials in Equation (3.2.3) might exhibit a unit root. This is the chief motivation of the so-called Integrated GARCH model.

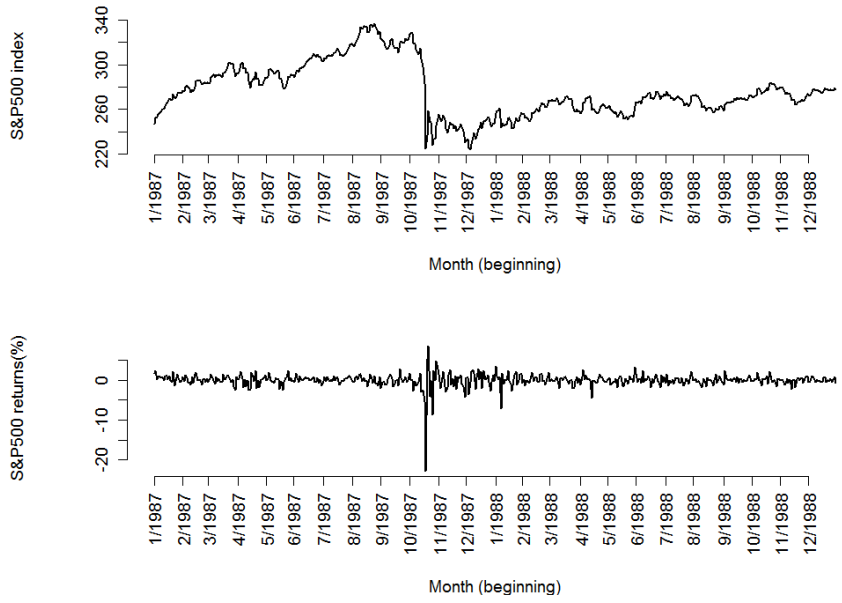


Figure C.1 S&P500 index and returns near the October 19, 1987 market crash. Data acquired from `http://finance.yahoo.com`, for the period from 01 Jan 1987 to 31 Dec 1988.

Since the persistence parameter $\phi = \alpha + \beta$ is very close to 1 as suggested by many empirical estimates, it seems only logical to assume a unit root exist for the conditional volatility process. We can test this null hypothesis against its alternative ($\phi < 1$), which is analogous to the assumption that the process is covariance stationary and has a non-negative finite variance. If the null hypothesis is not rejected, then the impact of returns shocks persist indefinitely. To some extent, this helps explain the long-memory behaviour of returns as implied by the long run significant autocorrelation of the volatility process. The extreme persistence is incorporated into the Integrated GARCH (IGARCH) model first considered by Engle and Bollerslev (1986), then further examined by Nelson (1990):

$$\sigma_t^2 = \omega + \alpha(u_{t-1}^2) + (1 - \alpha)\sigma_{t-1}^2 \tag{C.1.1}$$

As we can see the integrated GARCH process is no longer covariance stationary (as its unconditional variance tends to infinity). However Nelson (1990) proved that it is strictly stationary and therefore still capable of demonstrating highly persistent correlation structure of absolute returns. In other words the IGARCH is analogous to an ARIMA model. We note that the unit root hypothesis is generally not valid in exchange rate studies, and evidence of such a property in equity markets is mixed (See Pagan and Schwert (1990b) for supportive evidence and Blair, Poon, and Taylor (2002) for contrasting evidence of unit root existence).

Reider (2009) commented that since the impact of returns shocks are persistent, the volatility process follows a 'random walk' with zero 'drift' ($\omega = 0$). Also the unconditional variance is zero. Therefore the best l-period ahead prediction of volatility is the current level of variance:

$$\hat{\sigma}_{t+l}^2 = \sigma^2 + (\alpha + \beta)^l(\sigma_t^2 - \sigma^2) = \sigma_t^2$$

C.2 Fractionally integrated GARCH

The IGARCH specification is very extreme in the sense of allowing an infinite persistence, compared to the exponentially decaying autocorrelation of a standard GARCH process. Bollerslev and Mikkelsen (1996) pointed out that this may be because Nelson (1990)'s test for a unit root is too restrictive. To make up for this, these authors complement the alternative hypothesis, stating that volatility process is not necessarily restricted to be covariance stationary but could also be a "fractionally" integrated process, rather than a "fully" integrated one. This new model allows for a hyperbolically decaying autocorrelation while forfeiting the assumption of unit root behaviour. In other words, it strikes a "middle ground"

between GARCH and IGARCH. The model is then aptly named Fractionally integrated GARCH (FIGARCH).

The FIGARCH is analogous to the ARFIMA model discussed in Chapter 2. In the ARFIMA model, the long-memory operator (or fractional differencing operator) is applied to the unconditional mean (which is a constant) of a stationary process $\{X_t\}$, whereas applying it to the squared errors, u_t^2, gives us the equivalent of the ARFIMA (p,d,q) model for the conditional variance, i.e. the FIGARCH (p,d,q). Specifically, this model is defined as:

$$\begin{aligned} &\phi(L)(1-L)^d\, u_t^2 = \sigma_0 + [1-\beta(L)]\nu_t \\ &\nu_t = u_t^2 - \sigma_t^2 \\ &\sigma_t^2 = \omega[1-\beta(L)]^{-1} + \{1-[1-\beta(L)]^{-1}\phi(L)(1-L)^d\}u_t^2 \end{aligned} \tag{C.2.1}$$

where $\phi(L)$ and $\beta(L)$ play identical roles as the polynomials $a(L)$ and $b(L)$ do in the ARFIMA model (see subsection 2.1.3). The ν_t process can be thought of as the innovations for conditional variance. Bollerslev and Mikkelsen (1996) show that FIGARCH (p,d,q) is not covariance stationary. However, with $0 < d < 1$, the model is strictly stationary, now that the roots of $\phi(L)$ and $[1-\beta(L)]$ are no longer restricted to lie within the unit circle. For the model to be well-defined, all the coefficients in the infinite ARCH representation in Equation (C.2.1) must be non-negative.

Analogous to the ARFIMA process, for FIGARCH the autocorrelation of squared returns is asymptotically decaying at a slower rate than the GARCH process when $0 < d < 0.5$, thus accounting for the long-memory behaviour of the volatility of financial returns. The bulky estimation of this model is performed thanks to the versatile G@RCH package incorporated in the OXMetrics software (Laurent and Peters (2002)).

C.3 Asymmetric GARCH

Aside from the models discussed above, which are capable of capturing various properties of financial returns' volatility, another branch of models was designed for modelling the asymmetric effect of equity price changes on volatility. The success of original GARCH lies in its ability to model (time-varying) volatility clusters via the changes in residual returns. However, it only accounts for the effect of returns shocks' magnitude (where large shocks are followed by increases in volatility) while ignoring the effect of their signs (Awartani and Corradi (2005)).

The basic motivation for these models is empirical evidence of negative correlation between stock returns and volatility. In theory, negative returns (de-

creases in equity value) raise the financial leverage and make it riskier to hold stocks (Awartani and Corradi (2005)) hence the term "leverage effect" (See Black (1976), Christie (1982) and Pagan and Schwert (1990a) for early discussions). Engle and Ng (1993) examine volatility shocks, which are considered a product of a collection of news at time t. The impact of (unexpected) news on price is reflected by the sign and size of the unexpected returns u_{t-1}. The leverage effect generally describes the fact that bad news (reflected in price drops) tend to increases conditional volatility by a greater amount than good news does.

In addition, Bollerslev and Wright (2000) discussed the relevant literature regarding the documented negative relationship between unexpected returns and future volatility. In particular, the authors investigate two related theories, i.e. leverage effect and volatility feedback, which place opposite propositions as to what is the cause of this relationship. Previous studies examining low-frequency data suggests their respective impact is indistinguishable. However, by using a high frequency, aggregated index dataset, these authors' analysis indicates evidence in clear favour of persistent leverage effect, as well as yielding superior estimates of asymmetric impact of returns shocks on volatility. Given these insights, here we examine two of the simplest models in the asymmetric GARCH class in hope of capturing the leverage effect.

EGARCH : Nelson (1991) developed a new model with an ARMA (p,q) parametrization for the logarithm of σ_t. In particular, this model incorporates an asymmetric framework of the standardized residual returns $z_{t-1} = u_{t-1}/\sigma_{t-1}$:

$$\ln \sigma_t^2 = \omega + [1 - \varphi(L)]^{-1} \, [1 + \psi(L)] g(z_{t-1}) \tag{C.3.1}$$

Equivalently,

$$\begin{aligned} \ln \sigma_t^2 &= \omega + \alpha[g(z_{t-1})] + \beta \ln \sigma_{t-1}^2 \\ g(z_{t-1}) &= \vartheta z_{t-1} + \gamma\Big(|z_{t-1}| - E[|z_{t-1}|]\Big) \end{aligned} \tag{C.3.2}$$

Elaborating on this model, Taylor (2005) (Chapter 10) notes two important points:

- When returns are assumed to have a conditional Gaussian distribution then $E[|z_{t-1}|] = \sqrt{2/\pi}$ and the term $g(z_{t-1})$ follows a zero mean i.i.d process.
- Function $g(z_{t-1})$ has slope of $\vartheta - \gamma$ when $z_{t-1} < 0$ and of $\vartheta + \gamma$ otherwise. Therefore we expect $|\vartheta - \gamma| > |\vartheta + \gamma|$ when the assumption of asymmetric impact of returns shock holds. In other words ϑ and γ are expected to have opposite sign. Model estimates from Nelson (1991) for U.S. market indexes

show $\vartheta < 0$ and $\gamma > 0$. This is in line with Engle and Ng (1993)'s findings and our estimates (which are shown in section 3.3, Chapter 3).

GJR-GARCH : Glosten et al. (1993) model asymmetric volatility effect by adding a dummy variable, S^-_{t-1}, accounting for both positive and negative shocks in returns. The model was named after its creators - the GJR-GARCH:

$$S^-_{t-1} = \begin{cases} 1 & \text{if} \quad u_{t-1} < 0 \\ 0 & \text{if} \quad u_{t-1} \geq 0 \end{cases} \tag{C.3.3}$$

$$\sigma^2_t = \omega + (\alpha + \alpha^* S^-_{t-1}) u^2_{t-i} + \beta \sigma^2_{t-1} \tag{C.3.4}$$

Explicitly, a positive and significant value of α^* indicates greater impact of negative innovations compared to positive ones.

News impact curve Concurrent with the development of the asymmetric GARCH models discussed above, Engle and Ng (1993) proposed the use of a visual inspection of the implied relationship between past return shocks and present volatility. The method was first introduced by Pagan and Schwert (1990a). Providing that lagged conditional variance is evaluated by its unconditional level σ^2, the news impact curve of EGARCH (1,1) is given by the following equation:

$$N(u_{t-1}) = \sigma_t = \begin{cases} A \exp\left[\left(\dfrac{\vartheta + \gamma}{\sigma}\right) u_{t-1}\right] & \text{for} \quad u_{t-1} \geq 0 \\ \\ A \exp\left[\left(\dfrac{\vartheta - \gamma}{\sigma}\right) u_{t-1}\right] & \text{for} \quad u_{t-1} < 0 \end{cases} \tag{C.3.5}$$

where $A \equiv \sigma^{2\beta} \exp\left[\omega - \gamma\sqrt{2/\pi}\,\right]$

From a visual perspective, this curve reaches its minimum value $N_{\min}(u_{t-1}) = A$ when $u_{t-1} = 0$, and increases exponentially in both directions with different parameters (Engle and Ng (1993)). Because of asymmetric effect of shocks, we expect to see that the curve is steeper on the left-hand side of the vertical axis (when $u_{t-1} < 0$) than on the other side, and increasing on both sides. As in our discussion following model definition, the asymmetric curve "skews" to the left when $\vartheta < 0$ and $\gamma > 0$. The evidence of asymmetric positive impact of both negative and positive shocks is in line with the findings in Nelson (1991), but in partial contrast to that of Glosten et al. (1993), who conclude that whereas

negative shocks lead to increasing conditional variance, the expected impact of a positive unexpected return is negative. It is worth noting that these authors examine data at different frequencies: while Engle and Ng (1993) and Nelson (1991) use daily data, Glosten et al. (1993) use monthly data. In addition, the news impact curve of GJR-GARCH is specified via:

$$N(u_{t-1}) = \sigma_t = \begin{cases} A + \alpha\ u_{t-1}^2 & \text{for} \quad u_{t-1} \geq 0 \\ A + (\alpha + \alpha^*) u_{t-1}^2 & \text{for} \quad u_{t-1} < 0 \end{cases} \tag{C.3.6}$$

$$\text{where} \quad A \equiv \omega + \beta\ \sigma^2$$

Plotting news impact curves is a simple method to illustrate the difference among various volatility models. Engle and Ng (1993) also pointed out that while various asymmetric models may have different ways of capturing asymmetry via the news impact curve, they are all distinctive from the symmetric curve of standard GARCH which is centered at $u_{t-1} = 0$. One of their major propositions is that this difference is very important in determining volatility predictions especially after intensive bad news arrivals in extreme events such as the 1987 crash. Consequently, it may lead to differences in the implication of capital asset pricing and option pricing models. The different asymmetric features as opposed to the symmetry of GARCH can be clearly seen from the news impact curves of EGARCH(1,1) and GJR-GARCH(1,1) specified for daily S&P500 index time series, which are featured in Figure C.2.

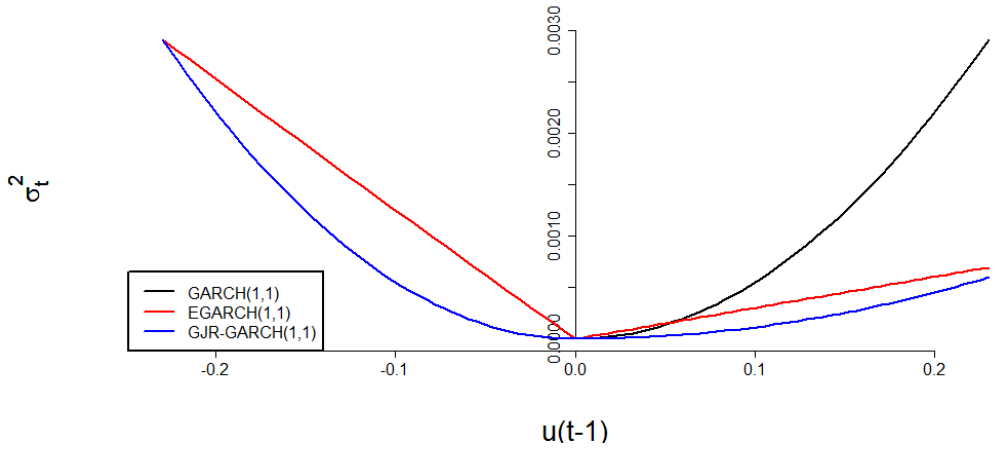


Figure C.2 News impact curves of different asymmetric GARCH models for daily S&P500 volatility. Data sample ranges from 03 Jan 1977 to 31 Jul 2013.

Diagnostic bias tests The bias tests' statistics are formulated in Engle and Ng (1993), and are incorporated in the R-package rugarch (Ghalanos (2013)) as well as in various econometric softwares. As an example, Table C.1 reports test outputs of the three models described above. All of the tests for EGARCH and GJR-GARCH indicate adequate account for asymmetric impacts of returns shocks. Only the test for negative size bias of GARCH yields significant coefficient, meaning our GARCH model does not distinguish between the impact of large and small negative shocks.

	Sign bias		**Negative size bias**		**Positive size bias**		**Joint Effect**	
	t-stat	p-value	t-stat	p-value	t-stat	p-value	F-stat	p-value
GARCH(1,1)	0.8072	0.4195	2.1627*	0.0306	0.7766	0.4374	5.5086	0.1381
EGARCH(1,1)	1.0171	0.3091	1.6279	0.1036	0.7085	0.4787	3.1583	0.3679
GJR-GARCH(1,1)	0.8329	0.4049	1.6218	0.1049	0.7348	0.4625	3.1991	0.3619

Table C.1 Diagnostic tests of different GARCH models for daily S&P500 index.
Data sample is from Dec 01, 1977 to 31 Jul, 2013.
(*) indicates significance at the 5% level.

Appendix D

The background of Citigroup

We choose to study one specific company in the financial service industry, Citigroup Inc., mainly for its central role in the recent financial crisis. Citigroup, Inc. (NYSE ticker symbol: 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 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 over-exaggerated 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.

Acknowledging the extent of actual name changes associated with the company's numerous branches, operations and subsidiaries (e.g. City bank, Citibank, Citicorp, Citi Holdings and their spin-offs, to list a few), we merely refer to the company as 'Citigroup' throughout the paper. The reason is that this name was largely maintained as the major branch name identifying the firm ever since the merger between giant bank Citicorp and the biggest insurance company in the U.S.A at the time - Travelers Group - in 1998. Following this $140b merger the firm became the world's largest financial service corporation, with asset value in excess of $1 trillion. It remained on the Dow-Jones Industrial Average index (DJIA) until mid 2009, following an announcement of significant government

ownership of 36% equity stake.

It is ironic that during the greater part of the 20th century, long before the turmoil of 2008 unfolded, Citigroup was largely credited with effective risk management as well as prudent lending practices. Then came the company's successive downturns, chief among which were (i) the steep losses from overseas lending and real estate lending realized in the early 1990s (as a result of the October 1987 market crash and in the wake of the U.S. savings and loan crisis); (ii) the non-synchronous operations of the merger partners after 1998, worsened by the busting of the Internet bubble in 2000 and the involvement with Enron' scandal in 2001; (iii) the over-investment in Collateralized debt obligation (CDO) and subprime mortgage loans during the first decade of this century. Despite being accused by many of being a major catalyst of the 2008 crisis due to its 'shadow banking' businesses and its deviation from previously prudential policies, this 'too big to fail' firm was the receiver of a large rescue package from the U.S. Federal Reserve in January 2009 (Morgenson (2011)) on the condition of tightening management. In the wake of much needed reconstruction, it had to get rid of many of its swollen under-performing and non-core businesses. It appears that Citigroup has succeeded in rebuilding its reputation, as by the end of 2010 it paid back the government bailout and started to realize profits while refocusing on its traditional banking business. In the aftermath of the GFC, the company has emerged as the second best performing American bank, after Bank of America.

Implications of stock splits and dividend adjustments Because of the extraordinary magnitude of various market crashes occurring between 1977 and 2013 as well as the frequent repurchases and restructuring, especially in the 1990s (as can be seen in Figure 4.1), it is difficult to obtain a comprehensive grasp of the fluctuations of the daily time series from an aggregated point of view, especially for the returns series. Therefore, to have a better visual impression of how the movements of these time series interact with each other we separately look at three consecutive periods with roughly equal length of 12 years: 1977-1989, 1990-2001, 2002-2013 and plot the two price series on the same scale. These 'rescaled' subperiods are displayed in Figure D.1. As can be seen, there are several important implications that can be derived from these plots:

- When being plotted on the same scale, the original close-to-close price series and its adjusted counterpart exhibit considerable difference in terms of magnitude, save for the very end of sub-period three. One can argue that the differences may only be a matter of pure technical adjustments and restructuring, as the pattern of original price closely resembles that of

adjusted price if we ignore the scaling factor. When constructing prices adjusted for splits and dividends, the most recent observations of closing prices are kept unchanged and form the base to " work backwards" to get the adjusted prices. Apart from stock splits and dividend considerations, the two price series are proportional and thus lead to identical returns value save for some abnormal realizations. In addition, the adjusted price series seems to be subject to rounding error from keeping insufficient significant figures (an issue which we do not further examine).

- For the first 12 years, original price is consistently greater than adjusted price. However most of the differences are caused by pure adjustments, as we can see the similarity in the movement of the two price series. Consequently, their respective returns are largely identical and it can be said that most unadjusted returns in this period reflect 'genuine' firm performance.
  This is not always true, however: the large stock split at the beginning of 1987 (highlighted by the red oval) is too remarkable for the argument of a genuine return to be valid. Indeed, this fall was 'washed-out' in the adjusted price returns. Specifically, the original returns realized a decrease of nearly 70% while at the same time adjusted returns observed no significant volatility. It can be concluded that the fall is merely an artefact of some restructuring, presumably a stock split after the large momentum built up before 1987.
  An additional significant fall is observed at the end of 1987 (highlighted by the blue oval). This time it is reflected in both original and adjusted price. Consequently, both returns series realized a fall of about 25%. This clearly marks the effect of the market crash in October 1987, and is a genuine bearish move of the stock rather than a mere technical adjustment.

- Similar arguments can be made to justify the differences between original returns and adjusted returns observed in the period 1990-2001. There were several stock splits/restructuring during this period. The major split right after the merger in June 1998 was reflected in both returns series. In fact this could be considered a stock swap, with Travelers Group issuing 2.5 new Citigroup shares for each Citicorp share it purchased in a record $70b deal (Martin (1998)). In addition, there seems to be minor stock splits in early 1999 and late 2000 which were not reflected in adjusted price series.
  The adjustments are all reflected by the difference in returns series. The lower bound of adjusted returns is roughly -10%, whereas that of original returns exceeds -40% (marked by the red ovals). Obviously there are

several artificial negative returns that can be attributed to frequent stock adjustments over the 12 years.

- The final session of the most recent period of two price series, from early 2011 to mid 2013, are in fact identical. For this period there is only one relatively dramatic restructuring recorded at early 2011. This was a result of a big-scale 'reversed' stock split in May, following the repayment of tax-payer funded aid. Indeed, the split is that 10 old shares merged to one new share, leading to price increase from around $5 to $45 overnight. Working backwards to compute the adjusted price, it inevitably assumes a boost in value by a factor of 10: for example, compared to a closing price of $60 in early 2007, the corresponding adjusted price is about ten times as great.
  2010 was the first profitable year of Citigroup since 2007. A $10.6b in net profit was reported, compared with a $1.6b loss in 2009 (Dennis (2010)). Finally emerging out of insolvency after two years of crisis, the company is well on track to profitability and prudential operations in the present days and the future.

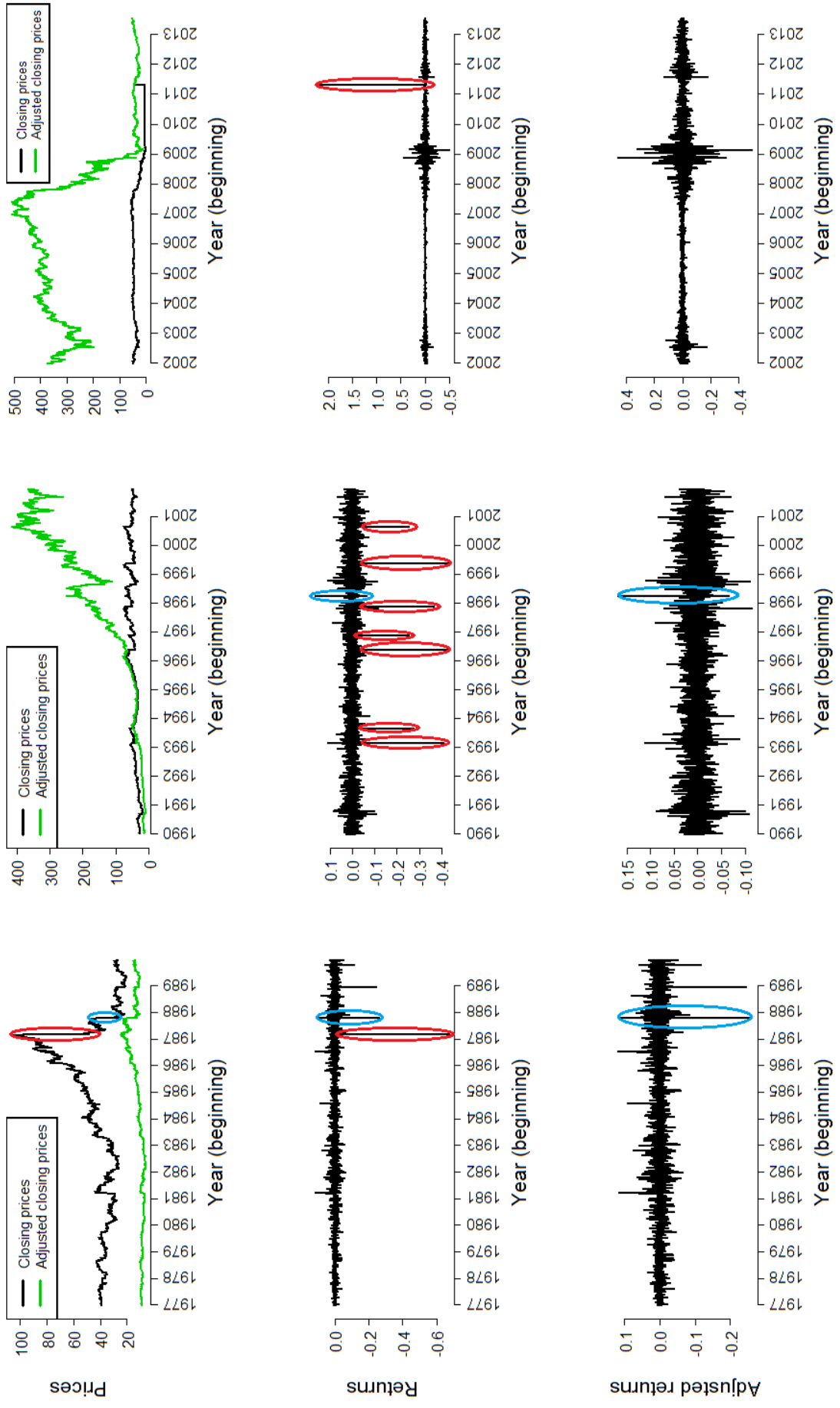


Figure D.1 Time series plot of Citigroup daily data for three consecutive periods: 1977-1989 (left), 1990-2001 (middle) and 2002-2013 (right). Red ovals indicate "artificial" returns (which disappear from adjusted returns) while blue ovals indicate "genuine" ones (which are retained in adjusted returns).