



City Research Online

City St George's, University of London

Citation: Hallam, Mark (2013). Essays on the applications of distributional scaling in finance: Estimation, forecasting and inference. (Unpublished Doctoral thesis, City University London)

This is the unspecified version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://city-ref-test.eprints-hosting.org/id/eprint/3042/>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Essays on The Applications of Distributional Scaling in
Finance: Estimation, Forecasting & Inference

Mark Hallam

A Thesis Submitted for the Degree of Doctor of Philosophy in Economics

Department of Economics, City University London

November 2013

Abstract

This thesis addresses some of the current gaps in the literature on unifractal and multifractal processes in finance, through a combination of empirical and theoretical contributions spanning the key problems of estimation, forecasting and inference.

In Chapter 2 a new method is proposed for producing density forecasts for daily financial returns from high-frequency intraday data, under the assumption that the return process possesses distributional scaling properties consistent with that of a unifractal process. In contrast to previous methods using intraday data to estimate and forecast daily return densities, the approach presented preserves information about both the sign and magnitude of the intraday returns and allows nonparametric specifications to be employed for the distribution of daily returns.

The density forecasting performance of the method is shown to be competitive with existing methods based on intraday and daily returns for exchange rate and equity index data, particularly for shorter in-sample periods and during periods of high return volatility. However, as expected the performance of the method is stronger for return series with distributional scaling properties close to the unifractal scaling required by the method and poorer, though still competitive, for time series that exhibit larger deviations from unifractality.

In response to the apparent limitations of the method proposed in Chapter 2, Chapter 3 develops an equivalent density forecasting method under the assumption that the return process belongs to the more general class of multifractal processes, thus permitting more flexible scaling behaviour than in Chapter 2. Whilst these distributional scaling laws are more problematic to apply in practice than those of Chapter 2, both the daily return variance and kurtosis can be estimated from the intraday data, providing additional flexibility over existing realised volatility based methods. The predictive ability

of this alternative multifractal density forecasting approach is found to be competitive with existing density forecasting methods for both exchange rate and equity index data, but is outperformed by the unifractal approach of Chapter 2 for equity index data.

Finally in Chapter 4, a formal testing framework is developed for determining whether a given sample of data is most consistent with a unifractal or multifractal data generating process. The testing methodology begins by proposing a set of possible statistics for testing the null hypothesis of unifractality against the alternative of multifractality, but due to the specific characteristics of the testing environment the distributions of the proposed test statistics are non-standard and the relevant rates of convergence are unknown. It is then shown that these difficulties can be overcome and test statistic distributions obtained using an appropriate model-based bootstrap resampling scheme.

A series of Monte Carlo exercises demonstrate that the testing procedure possesses good empirical size and power properties in wide range of situations, being robust against various forms of multifractality under the alternative. Good performance for sample sizes that would be considered as small in the multifractality literature also confirms the suitability of the methodology for the study of both local and global scaling properties. This is demonstrated in an empirical exercise in which the testing methodology is applied to study the local scaling properties of the intraday dataset used in previous chapters.

Acknowledgements

Firstly, I want to thank Jose Olmo for his supervision and guidance over the past four years, not only related to the writing of this thesis, but also on academic issues more generally.

From City University I must thank everyone who has provided comments and suggestions at various points that have influenced this work. This includes, but is not limited to, Michael Ben-Gad, Giulia Iori, Saqib Jafarey, Gabriel Montes-Rojas and Yang Zu. From City I also want to thank Andy Denis, who has not been directly involved in the creation of this thesis, but whose attitude towards learning and teaching has been a great inspiration for me.

From elsewhere I should also thank Peter Burridge from the University of York for his supervision during my Masters programme and for encouraging me to move from York to London for my doctoral studies. I am also most grateful to Kamil Yilmaz and others at Koç University in Istanbul for making my time spent there over the past year as productive and comfortable as it has been.

Finally, I want to thank all of my family spread across two continents; I would not be where I am or who I am without their love and support and I certainly could not have completed this thesis without their help.

Contents

1	Introduction	6
2	Density Forecasts of Daily Financial Returns from Intraday Data I: A Unifractal Approach	13
2.1	Introduction	14
2.2	Theory & Application of Unifractal Processes	17
2.2.1	A Review of Unifractal and Multifractal Processes	17
2.2.2	Estimating Daily Return Densities from Intraday Data	20
2.3	Estimating the Density of Daily Returns from Intraday Data	22
2.3.1	Estimation of the Self-Affinity Index	23
2.3.2	Estimation of Daily Return Densities	25
2.4	Forecasting the Density of Daily Returns	28
2.4.1	The Forecasting Environment	29
2.4.2	Dynamic Structure for Density Forecasting: Parametric Case . . .	30
2.4.3	Dynamic Structure for Density Forecasting: Nonparametric Case .	31
2.4.4	Estimating Autoregressive Parameter Values	33
2.5	Empirical Application	36
2.5.1	Data	37
2.5.2	Benchmark Density Forecasting Models	41
2.5.3	Method for Density Forecast Comparison	42

2.5.4	Empirical Results	44
2.6	Conclusion	52
3	Density Forecasts of Daily Financial Returns from Intraday Data II:	
	A Multifractal Approach	56
3.1	Introduction	57
3.2	Unifractal & Multifractal Processes	59
3.2.1	A Review of Unifractal and Multifractal Processes	59
3.2.2	Application of the Distributional Scaling Laws	62
3.3	Estimating and Forecasting the Moments of Daily Returns from Intraday Data	65
3.3.1	Estimation of the Multifractal Scaling Function	65
3.3.2	Forecasting the Moments of Daily Returns	68
3.3.3	Producing Density Forecasts from Point Forecasts of Moments	72
3.3.4	Two Sources of Multifractality and an Extension to the Method	75
3.4	Empirical Exercise	77
3.4.1	Data	77
3.4.2	Benchmark Density Forecasting Models	79
3.4.3	Methods for Density Forecast Comparison	81
3.4.4	Empirical Results	85
3.5	Conclusion	95
4	A Statistical Test for Unifractality Versus Multifractality	99
4.1	Introduction	100
4.2	A Review of Unifractal and Multifractal Processes	104
4.2.1	Theoretical Scaling Properties	105
4.2.2	Estimation of Scaling Properties	110
4.2.3	Previous Graphical Tests for Scaling & Spurious Multifractality	117

4.3	Testing Methodology	121
4.3.1	A Summary of the Testing Problem	121
4.3.2	Hypothesis Tests of Unifractality Versus Multifractality	122
4.4	Resampling Methods for Obtaining Test Statistic Distributions	130
4.4.1	Issues Affecting the Choice of Resampling Method	130
4.4.2	Model-based Bootstrap Resampling	132
4.4.3	Alternative Resampling Approaches	137
4.5	Monte Carlo Exercises	139
4.5.1	Monte Carlo Methodology	139
4.5.2	Core Results & the Effects of Sample Size	143
4.5.3	Robustness to Changes in the Estimation Method	145
4.5.4	Choice of Process Under the Null and Alternative	148
4.6	Empirical Exercise	153
4.7	Conclusion	157
5	Concluding Remarks	165

List of Figures

2.1	Estimated unifractal and multifractal scaling functions	40
2.2	Period-by-period CRPS differential and daily realised volatility.	53
3.1	Estimated unifractal and multifractal scaling functions	80
3.2	Multifractal moment estimates from shuffled and ordered intraday data .	91
4.1	Theoretical values of the scaling function and generalised Hurst exponent for a multifractal random walk process	108
4.2	Theoretical values of the scaling function and generalised Hurst exponent for a generic unifractal processes	110
4.3	Estimated values of the scaling function and generalised Hurst exponent for financial data	160
4.4	Estimates of the scaling function and generalised Hurst exponent for unifractal fractional Gaussian noise realisations	161
4.5	Empirical power versus multifractality strength (λ^2) for dH_{inf} and dH_{avg} statistics	162
4.6	P-values for tests of unifractality versus multifractality applied to consec- utive non-overlapping sub-periods of 5-minute EUR/USD data	163
4.7	P-values for tests of unifractality versus multifractality applied to consec- utive non-overlapping sub-periods of 5-minute S&P500 data	164

List of Tables

2.1	Density forecast comparison against AR-RV(5) benchmark	46
2.2	Density forecast comparison against GARCH(1,1) benchmark	48
3.1	Sensitivity of predictive ability to changes in estimation window length . .	86
3.2	Out-of-sample density forecast comparison using CRPS-based test statistic	87
3.3	Out-of-sample density forecast comparison using CRPS-based test statis- tic - shuffled data with 1000 repetitions used for multifractal approach . .	92
3.4	Certainty equivalent returns from portfolio allocation exercise	94
4.1	Empirical size and power for default testing methodology	145
4.2	Empirical size and power for different choices of $\mathbf{q}$	146
4.3	Empirical size and power comparison between MF-DFA & MF-CMA es- timators	148
4.4	Effect of the strength of multifractality on empirical power	149
4.5	Empirical power against Markov-switching multifractal alternative	151
4.6	Empirical size and power with linearly dependent series	152

Chapter 1

Introduction

The work presented in this thesis centres around the theory and applications of unifractal and multifractal processes in the context of finance. Whilst much of the work on these processes originated in physics, where they are used in the study of turbulence and other complex systems, they have also been found to be highly relevant for the modelling of financial time series. On an intuitive level, unifractal and multifractal processes exhibit a form of scale invariance or distributional scaling, such that the statistical properties of the process observed at one timescale are formally linked to those at other timescales according to some theoretical scaling laws.

Using the numerous methods available for estimating the distributional scaling properties of a given sample of data, a large number of empirical studies have found scaling consistent with these processes in the return series for a wide range of assets including equities, exchange rates, fixed income securities and commodities. These empirical studies constitute a substantial proportion of the total literature on fractal processes in finance and are too numerous to list completely, however some of the more notable examples include Schmitt et al. (1999), Calvet and Fisher (2002), Fillol (2003), Xu and Gençay (2003), Di Matteo et al. (2005), Selçuk and Gençay (2006), Di Matteo (2007) and Onali and Goddard (2009).

The empirical evidence in support of distributional scaling in asset returns has in turn motivated the development of various theoretical processes for modelling financial returns, which reproduce both the relevant distributional scaling properties and other key characteristics of financial returns, such as volatility clustering. These include the multifractal model of asset returns of Mandelbrot et al. (1997) and Calvet and Fisher (2002), the random walk of Bacry et al. (2001) and the Markov-switching multifractal of Calvet and Fisher (2004). A small number of studies have even extended the application of these multifractal processes beyond the context of financial returns to the problem of modelling other financial time series, such as the Markov-switching multifractal model for intra-trade duration recently proposed by Chen et al. (2013).

Overall, much has been achieved towards understanding how and where these processes can be appropriately applied in financial applications and developing suitable methods and models to do so. However, there are still noticeable gaps in the existing literature requiring further study, with the current work aiming to address some of these through a combination of theoretical and empirical contributions.

The first major shortcoming in the literature is the relative lack of effort made to develop methods for actually applying the unique properties of unifractal and multifractal processes to common financial problems, such as risk management, portfolio allocation or asset pricing. The large number of empirical studies have provided substantial evidence that financial returns are consistent with these processes, but are almost exclusively exploratory in nature and devoid of any theoretical content or attempts to apply the observed properties in any meaningful way. In the more theoretical branch of the literature, methods have been developed for some of the specific unifractal and multifractal processes proposed that could be used for practical financial problems, such as the frameworks for forecasting return volatility within the Markov-switching multifractal and multifractal random walk models proposed by Calvet and Fisher (2004) and Duchon et al. (2010) respectively. However, such examples are not common and elsewhere in the literature very little effort has been made to fully exploit the properties of these processes more generally.

Given that the distributional scaling laws satisfied by unifractal and multifractal processes provide a formal link between the properties of the process at different timescales, an obvious financial application of these processes is to relate the properties of intraday returns to those of daily or lower frequency returns; this would allow intraday information to be incorporated into estimates and forecasts for returns at lower frequencies. The large existing literature on realised volatility and related measures has already demonstrated the potential gains in estimation and forecasting performance that can be obtained from the effective incorporation of intraday data into models for lower frequency returns. The

distributional scaling properties of fractal processes can in principle be employed in an equivalent manner and establishing the feasibility and potential gains of such methods is the primary focus of Chapters 2 and 3 of this thesis.

A secondary contribution of these two chapters is the move to a dynamic estimation context, allowing the structure of distributional scaling to change over time. The vast majority of the existing literature does not allow for such a possibility, imposing a static estimation environment with the scaling properties of the return process assumed to be fixed for the complete sample period. By allowing the distributional scaling properties to be time-varying, additional flexibility is introduced into the framework and this leads naturally to the final aim of producing forecasts to be employed for financial problems such as portfolio allocation or risk management.

Initially in Chapter 2 a new method is proposed for producing semiparametric density forecasts for daily financial returns from high-frequency intraday data, under the assumption that the return process possesses distributional scaling properties consistent with that of the class of unifractal processes; this imposes a more restrictive form of distributional scaling than that in the multifractal context, but leads to simple and flexible implementation of the relevant scaling laws that is not possible in the more general multifractal case.

In contrast to previous methods using intraday data to estimate and forecast daily return densities based on realised volatility measures, the unifractal approach presented preserves information about both the sign and magnitude of the intraday returns. Furthermore, the unifractal approach allows nonparametric specifications to be employed for the distribution of daily returns, thus avoiding the potential difficulties encountered in selecting a suitable parametric model for asset returns and allowing the intraday returns to influence all aspects of the daily return density. The performance of this proposed unifractal method is assessed in an empirical exercise based on intraday data for a set of key exchange rate and equity index series. The out-of-sample density forecasting per-

formance will be compared against that of existing forecasting methods based on both intraday and daily returns to establish the potential gains the additional theoretical flexibility provided by the new unifractal density forecasting method.

In response to the potentially limited applicability of the method proposed in Chapter 2 for return series that deviate from the required assumption of unifractal distributional scaling, Chapter 3 develops an equivalent density forecasting method under the assumption that the return process belongs to the more general class of multifractal processes. This assumption permits more flexible distributional scaling behaviour across timescales than the unifractal processes of Chapter 2, but the distributional scaling laws are more problematic to apply in practice. Most notably nonparametric specifications can no longer be used for the daily return density, however both the daily return variance and kurtosis can be estimated from the intraday data, thus providing additional flexibility in principle over existing realised volatility based methods.

The predictive ability of this alternative multifractal density forecasting approach is again compared to that of existing methods based on daily and intraday data. Furthermore, the criteria for comparing forecast performance is expanded from the purely statistical loss function of Chapter 2, to include an economic loss function in the form of a portfolio allocation exercise between a risky and risk free asset. Given that the multifractal forecasting approach of Chapter 3 allows for more flexible distributional scaling than the unifractal approach of Chapter 2, but imposes more restrictions on the implementation of these scaling laws, it is not clear a priori which of the two methods will provide superior forecasting performance. Therefore, the relative predictive ability of the proposed unifractal and multifractal methods is compared directly during the empirical exercise in order to establish which of these theoretical strengths proves to be most advantageous in practice.

The final gap in the literature that the current work aims to address in Chapter 4 is more theoretical in nature and is the lack of a formal statistical test for determining

whether the distributional scaling properties of a given sample of data are more consistent with a unifractal or multifractal process. As briefly mentioned above and discussed in detail later, the theoretical properties of the two types of process differ, with multifractal processes allowing for more flexible distributional scaling across timescales. The cost of this additional theoretical flexibility is that problems such as parameter estimation, forecasting and simulation are all more complex and computationally demanding than in the simpler unifractal case.

Given this tradeoff between theoretical flexibility and practical complexity for these two classes of process, the question of whether a specific sample of data is most consistent with a unifractal or multifractal process is clearly an important issue. This is true when applying these processes more generally, but especially so for the methods proposed in the earlier chapters of the current work, since the type of process dictates which of the two estimation and forecasting methods is appropriate for a given dataset.

Despite the obvious importance of this issue, no formal statistical test has been developed for this purpose that is generally applicable, with previous empirical studies relying on informal graphical procedures when testing for multifractal or unifractal scaling in financial data. Whilst this graphical approach is expected to perform acceptably in most situations when applied to large samples, it has been demonstrated in the literature that it may suggest multifractal scaling even for data generated by a purely unifractal process due to the problem of ‘spurious multifractality’ (see for example the work of Lux, 2004, Ludescher et al., 2011 or Schumann and Kantelhardt, 2011). The issue of spurious multifractality is particularly problematic in smaller samples, making the existing informal graphical testing approach particularly unsuitable for testing the local scaling properties of a process over shorter sub-periods of the total sample, which are arguably more relevant for the work of the earlier chapters than global scaling properties.

Therefore, the work of Chapter 4 develops a formal testing framework for determining whether a given sample of data is most consistent with a unifractal or multifractal

data generating process and avoids the problems associated with the existing testing methods. A set of possible statistics are proposed for testing the null hypothesis of unifractality against the alternative of multifractality for a given sample of data. Due to the specific characteristics of the testing environment and the complex theoretical properties of unifractal and multifractal processes, the distributions of the proposed test statistics are non-standard and the relevant rates of convergence are unknown. It is shown that these difficulties can be overcome through the use of an appropriate model-based bootstrap resampling scheme, allowing the distributions of the test statistics under the null of unifractality to be approximated in order to calculate critical values or p-values for the tests.

The size and power properties of the proposed testing procedure are evaluated in a series of Monte Carlo exercises using simulated unifractal and multifractal data. These exercises cover a wide range of situations in order to establish the robustness of the methodology to changes in either the implementation of the tests or the properties of the true data generating process under the null and alternative hypotheses. Finally, the testing methodology is applied to the dataset of intraday exchange rate and equity index data used in the previous chapters in order to examine both the global and local scaling properties of the data.

Chapter 2

Density Forecasts of Daily Financial Returns from Intraday Data I: A Unifractal Approach

2.1 Introduction

Over the past decade there has been a dramatic increase in the availability of intraday financial data, resulting in an extensive literature on the use of high-frequency data in financial econometrics. These data obviously allow for the study of financial market behaviour at intraday timescales, but they also contain potentially valuable information for longer timescales, which are arguably of more interest for most market participants. As a result, there have been efforts to incorporate intraday data into the modelling and forecasting of financial variables at daily or even lower frequencies.

The most notable example is provided by the large literature on realised volatility, a concept that was first properly formalised by Andersen et al. (2001). Realised volatility and related measures allow the unobservable daily volatility to be estimated from intraday returns and it has been found (see for example Andersen et al., 2003) that such measures can provide significant improvements in the modelling and forecasting of daily return volatility compared to models using only daily data.

Whilst return volatility is undoubtedly a variable of substantial interest, there are situations in finance in which information concerning just the first two moments of the distribution of returns is not sufficient. Perhaps most obviously, risk management problems, such as the calculation of Value-at-Risk, require knowledge of particular quantiles of the return distribution. In addition, it has been shown that higher moments, such as skewness and kurtosis, are time varying and relevant for problems of portfolio allocation and asset pricing (see for example Harvey and Siddique, 2000, or Dittmar, 2002).

However, as noted by Žikeš (2009), the use of intraday data to model and forecast characteristics of daily returns beyond the first two moments has not yet received much attention. Notable exceptions include Andersen et al. (2003), Giot and Laurent (2004), Clements, Galvão, and Kim (2008) and Maheu and McCurdy (2010), all of which extend the use of realised volatility measures to either the quantiles or the density function of

daily returns. The methods used by these previous studies consist of two components. The first is a parametric time series model for volatility incorporating one or more realised volatility measures, which is used to model and produce point forecasts for daily volatility. The second component is a typically a parametric distributional assumption about daily returns, allowing density or quantile forecasts for daily returns to be produced from the point forecasts of daily realised volatility.

There are two potential weaknesses with this approach; firstly, the high-frequency data enter only through the realised volatility measures and so any information provided by the sign of the intraday returns is lost when they are squared. Furthermore, the intraday data can only directly influence the second moment of the daily return distribution. Secondly, with the exception of Clements, Galvão, and Kim (2008) who also consider an empirical distribution for returns, these previous studies require a specific parametric form to be chosen for the distribution of daily returns; choosing the most appropriate parametric distribution for financial returns is difficult, particularly in a dynamic context, and the density forecasts produced by misspecified parametric models will generally be misleading. The semiparametric quantile regression approach of Žikeš (2009) avoids the last of these problems, but only produces estimates for specific quantiles rather than the complete distribution or density.

As stated in Chapter 1, the current chapter proposes a new approach for the estimation and forecasting of daily return densities from intraday data based on the theory of unifractal processes, which is motivated by the above discussion. Under the assumption that the return process is unifractal, the distribution of returns at any pair of timescales is identical after rescaling by an appropriate factor; this factor can be estimated for a particular time series and used to rescale the intraday data for a given time period, such that they are equal in distribution to daily returns. The density of daily returns can then be estimated from these rescaled intraday observations.

The proposed method has two theoretical advantages compared to existing methods

based on realised volatility. Firstly, the daily return density is estimated directly from these rescaled intraday observations (rather than squared or absolute values), thus preserving information contained in both the magnitude and sign of the intraday returns. Secondly, because a large sample of rescaled intraday observations are obtained for each trading day, it is possible to apply a range of estimators to these rescaled intraday returns to estimate the daily return density for a given trading day.

In particular, a nonparametric density estimation approach is proposed using a standard kernel density estimator, which allows the intraday data to influence all aspects of the daily return density without being complex to implement or computationally demanding. However, the use of nonparametric density estimators precludes the use of standard dynamic structures for forecasting and so a new method is proposed and developed that imposes a parametric dynamic structure directly on the time series of densities themselves, with the relevant parameters selected using concepts from the literature on density forecast combination.

The structure of the chapter is as follows: Section 2.2 presents the relevant theory on unifractal processes and describes how these results can be employed to estimate the density of daily returns from intraday return data. Section 2.3 details the chosen estimation methods for both the scaling factor used to rescale the intraday data and the density function of the rescaled intraday data. Section 2.4 explores the issue of producing density forecasts for daily returns from a time series of estimated return densities. Section 2.5 presents an empirical application comparing the density forecasting performance to existing methods using intraday data on equity indexes and exchange rates and finally Section 2.6 concludes.

2.2 Theory & Application of Unifractal Processes

In order to estimate the density of daily returns from intraday data a method is required for formally linking the characteristics of return distributions across different sampling frequencies. Instead of the realised volatility measures previously employed in the literature, the proposed method relies on results from the theory of self-affine or unifractal processes and the distributional scaling properties that these processes possess.

On an intuitive level, such stochastic processes exhibit some form of *scale invariance*, such that the behaviour of the process observed at one timescale is, after an appropriate transformation, identical in a statistical sense to that observed at another time scale. A large number of empirical studies have confirmed the existence of this type of distributional scaling behaviour in a wide range of financial time series and this has led to the development of several asset pricing models that explicitly reproduce this distributional scaling behaviour¹.

The current section begins with a brief summary of the theoretical properties of these processes (with more detailed treatments found in Calvet and Fisher, 2002 or Kantelhardt, 2009), before exploring how these properties can be employed to estimate the density of daily returns from intraday return data.

2.2.1 A Review of Unifractal and Multifractal Processes

The distributional scaling behaviour of a unifractal or self-affine process can be defined by a simple expression that links the distribution of the process at different sampling intervals. Formally, unifractal or self-affine processes can be defined in the following way:

Definition 2.1. A process is said to be *self-affine* or *unifractal* if for some $H > 0$, all $c \geq 0$ and all $t_1, t_2, \dots, t_k \geq 0$ it obeys the distributional scaling relationship

¹See again the list of references in Chapter 1 for examples of both empirical studies of distributional scaling in finance and asset pricing models based on these concepts.

$$\{X(ct_1), X(ct_2), \dots, X(ct_k)\} \stackrel{d}{=} \{c^H X(t_1), c^H X(t_2), \dots, c^H X(t_k)\} \quad (2.2.1)$$

which can be expressed more compactly as:

$$X(ct) \stackrel{d}{=} c^H [X(t)] \quad (2.2.2)$$

If the increments of the process are stationary, then the distributional scaling law of (2.2.2) also holds at the local level for the increments of the process²:

$$X(t + c\Delta t) - X(t) \stackrel{d}{=} c^H [X(t + \Delta t) - X(t)] \quad (2.2.3)$$

The parameter H is known as the self-affinity index and can be estimated for a specific time series of data using a variety of methods. For a self-affine process the self-affinity index coincides with the Hurst exponent that describes the long memory properties of the process and so the two terms are often used interchangeably in the context of unifractal processes. Indeed, for the more general multifractal processes considered below, the distributional scaling properties of the process are also closely linked to the serial dependence structure, however the relationship between the two is substantially more complex.

Common examples of unifractal processes in finance include the standard Brownian motion, for which $H = 1/2$, and also the more general fractional Brownian motion (and the corresponding increment series, the fractional Gaussian noise), for which H is constant but not constrained to be equal to $1/2$. In the current context, under the

²Note that (2.2.2) holds for any value of c and (2.2.3) implies that the increment process of X exhibits distributional scaling at any given sampling interval.

assumption of unifractality these scaling laws imply that the distribution of returns at different timescales or sampling intervals is identical after rescaling by a factor that depends on the characteristics of the particular return process (via H) and the difference between the two sampling intervals (via c).

One can also consider the more general class of multifractal processes, which allow for a more flexible relationship between distributions across different sampling frequencies. In the case of a multifractal process, equation (2.2.2) is generalised to:

$$X(ct) \stackrel{d}{=} c^{H(c)}[X(t)] \quad (2.2.4)$$

where the scaling factor c^H has been replaced by the more general function of c , $c^{H(c)}$. An alternative characterisation of scaling behaviour is often used in the multifractal case, where it can be shown (see for example Mandelbrot et al., 1997) that a stochastic process $X(t)$ with increments $X(t + \Delta t) - X(t)$ is multifractal if these increments are stationary and satisfy:

$$E[|X(t + \Delta t) - X(t)|^q] = c(q)(\Delta t)^{\tau(q)+1} \quad (2.2.5)$$

The function $\tau(q)$ in (2.2.5) is referred to as the *scaling function* and describes how different moments of the absolute increments of the process $X(t)$ scale with the sampling interval, Δt . It can be demonstrated (see Calvet and Fisher, 2002) that for a multifractal process the scaling function is strictly concave with intercept equal to -1. For a unifractal process (2.2.6) also holds, but the scaling function is linear and of the form $\tau(q) = Hq - 1$, where H is the same self-affinity index from equations (2.2.1) and (2.2.2). As with the self-affinity index for a unifractal process, the scaling function can be estimated for a particular time series using various methods (see Kantelhardt, 2009, for a survey of

several common estimators).

As previously stated in Chapter 1, the development of a parallel estimation and forecasting methodology under the more general assumption of multifractality is the focus of Chapter 3 and as such, a more detailed discussion of the class of multifractal processes is saved until the following chapter.

2.2.2 Estimating Daily Return Densities from Intraday Data

To proceed we require the following assumption:

A.1.1. The stochastic logarithmic price process, $X(t)$, is unifractal and has stationary increments $Y^\Delta(t)$, where $Y^\Delta(t) = X(t + \Delta) - X(t)$ is the return process over the interval Δ ³.

Whilst there are numerous empirical studies confirming the existence of distributional scaling behaviour in a wide range of financial assets (again, see footnote 2 for examples), these studies are typically concerned with the more general multifractal case, which includes the unifractal scaling of assumption A.1.1 as a special case. Whether a given sample of data is consistent with assumption A.1.1 is primarily an empirical issue and will be discussed further in the context of the current dataset during the empirical exercise.

For ease of exposition, the current subsection focuses on the specific example of estimating daily return densities from 5-minute intraday data. However, under assumption A.1.1 the distributional scaling laws of (2.2.2) and (2.2.3) hold for any pair of timescales and so the method could be used to estimate the density of returns for any given sampling interval from those at some higher frequency. However, if attention is restricted

³Whilst not directly relevant in the current context, this assumption may introduce the possibility of arbitrage opportunities due to the properties of some unifractal processes (see for example Bender et al., 2007).

to a specific pair of sampling intervals then a weaker condition than A.1.1 would be sufficient; it would then only be strictly necessary for (2.2.2) to hold for the sampling intervals of interest and not for all possible sampling intervals as is the case for a true unifractal process⁴.

Assume that a series of T 5-minute returns are observed for a financial asset over a given period and denote this set of intraday returns and the corresponding probability density function by:

$$\{Y_{I,t} : 1 \leq t \leq T\} \quad \text{and} \quad f(y_I)$$

where the I subscript is used to indicate returns at the intraday frequency. Under assumption A.1.1., by definition the return process must satisfy the distributional scaling laws of equation (2.2.3). If we denote the density of daily returns over the same period by $f(y_D)$, then in the context of the current example these scaling laws imply that:

$$f(y_D) = f(c^H y_I) \tag{2.2.6}$$

From (2.2.6) it can be seen that the density of daily returns is equal to the density of the 5-minute intraday returns, once these intraday returns have been appropriately rescaled by a factor consisting of two components; c and the self-affinity index, H . From (2.1) the value of c is determined solely by the relative lengths of the two sampling intervals; for a market with 24-hour trading, as is typical for FOREX, the appropriate value of c in the current example would be 288, since there are 288 5-minute returns observed over a 24-hour period. For a market with shorter trading hours, such as equity markets, c will take a smaller value.

⁴This weaker condition would allow for the distributional scaling relationship to change or break down at either very short or long sampling intervals that are outside the range of interest.

The self-affinity index, H , describes the relationship between the distributions (or probability densities) of the return process at different timescales and can be estimated from the intraday data using various estimators and the resulting estimate is denoted by $\hat{H}$. The intraday returns are then rescaled by the factor $c^{\hat{H}}$, with $c = 288$ as discussed above. The density of these rescaled 5-minute intraday returns can then be estimated and from (2.2.6), the resulting estimate can be viewed as an estimate of the density of daily returns over the same period. More formally:

$$\hat{f}(y_D) = \hat{f}(288^{\hat{H}} y_I) \quad (2.2.7)$$

where $\hat{f}(z)$ is used to denote an estimate of the probability density of the variable z . Therefore, under the assumption that the return process is unifractal, the density of daily returns for a given period can be estimated from the intraday returns observed over the same period using the distributional scaling laws of the previous section.

Whilst the above method for relating the distribution of intraday returns to that of returns at a lower frequency, through a rearrangement of the scaling law (2.2.3), may seem straightforward, it is important to note that this has not been proposed previously in the literature. Indeed as discussed in Chapter 1, outside of a small number of results derived for specific theoretical unifractal and multifractal processes, there has been no attempt to employ these scaling laws more generally for practical financial problems.

2.3 Estimating the Density of Daily Returns from Intraday Data

Estimation of the daily return density from the intraday data occurs in two stages: in the first stage, the self-affinity index is estimated from the intraday returns observed over a given time period and the resulting estimate is then used to rescale the intraday returns

as discussed in Section 2.2. In the second stage, the probability density function of these rescaled returns is estimated using the chosen density function estimator, providing an estimate of the daily return density for the same time period. Sections 2.3.1 and 2.3.2 will discuss in turn the methods employed for the first and second stages of estimation, respectively.

2.3.1 Estimation of the Self-Affinity Index

Numerous estimators are available for the self-affinity index of a unifractal process and several studies⁵ have demonstrated that the relative performance of these estimators can vary substantially in small samples. Due to their typically strong performance relative to other estimators and their suitability for the current dynamic estimation environment, attention was restricted to the class of estimators for the self-affinity index based on the local, rather than global, scaling properties of the process; these include the Detrended Fluctuation Analysis (DFA), Detrended Moving Average (DMA) and Centred Moving Average (CMA) estimators. The DMA estimator is detailed below, since it was typically found to produce superior density forecasting performance to the DFA for the current dataset, whilst the CMA estimator provided near identical performance to that of the chosen DMA method. Further details of these and other estimators can be found in the survey by Kantelhardt (2009)

Consistent with the previous notation, the increment series of the process of interest (the return series in the current context) is denoted by $\{y_t : t = 1, \dots, T\}$. The DMA estimate of the self-affinity index or Hurst exponent, H , is then obtained as follows:

1. Calculate the cumulative sum or ‘profile’ series as $x_t \equiv \sum_{t=1}^T y_t$. Note that in the current context, the profile series corresponds to the logarithmic price series and so this initial step can be avoided by simply beginning with the logarithmic price series directly.

⁵See for example Delignieres et al. (2006), Mielniczuk and Wojdyllo (2007) and Bashan et al. (2008).

2. A moving average of the profile series, denoted $\bar{x}_{s,t}$ is obtained for the window length s via a moving average filter of the form:

$$\bar{x}_{s,t} = \frac{1}{s} \sum_{j=0}^s x_{t-j}$$

for $t = s, \dots, T$. Note that the moving average filtered series is not defined for the first $s - 1$ observations of the original series.

3. The profile series x_t is detrended using the moving average trend series and the value of the sample fluctuation function (or generalised variance) for the current value of s , denoted $F_{T,s}^2$, is then obtained as:

$$F_{T,s}^2 = \frac{1}{T-s} \sum_{i=s}^T [x_i - \bar{x}_{s,i}]^2$$

The loss of some observations at beginning of the series can be avoided if desired by repeating the same process starting from the end of the series and averaging the two values of $F_{T,s}^2$ obtained from the forward and reversed series.

4. As $T \rightarrow \infty$, the sample fluctuation function defined above will converge to the population analogue, F_s , which for a unifractal process satisfies the following scaling relationship:

$$F_s \propto s^H \quad \text{or equivalently} \quad \log F_s = a + H \log s$$

where H is the self-affinity index or Hurst exponent. Steps 2-4 are repeated for different values of the segment size, s , between some minimum and maximum values s_{min} and s_{max} . The value of H can then be estimated from the slope of a linear fit of the logarithm of the sample fluctuation function $F_{T,s}$ against the logarithm of s , with linear regression typically used to obtain the estimate of H .

A brief note is required at this point regarding the issue of intraday seasonality. It is well known that intraday financial data can exhibit strong deterministic seasonal patterns that pose a problem for certain estimation methods. Because of this, the DMA estimator was applied both to the raw 5-minute returns and to intraday data seasonally adjusted using the method of Andersen et al. (2003). The resulting estimates of the self-affinity index from both cases were similar, as was the predictive ability of the resulting density forecasts with neither approach having a consistent advantage over the other. This is perhaps due to the fact that the chosen DMA estimator automatically performs local detrending across various window sizes in the process of estimating the self-affinity index, thus eliminating some or all of the seasonal patterns. Following this finding, it was decided to use the simple unadjusted intraday data for estimation of the self-affinity index, rather than the seasonally adjusted data.

2.3.2 Estimation of Daily Return Densities

Given that a large number of rescaled intraday returns can be obtained even for a single trading day, it is possible to apply a variety of estimators for the probability density function, including both parametric and nonparametric methods. Whilst the possibility of using nonparametric estimation methods is perhaps most interesting, one example of each class of estimator will be employed for the empirical analysis in order to investigate the potential gains from such a nonparametric approach. The current density function estimators are both intentionally simple, but could be replaced with more complex estimators without substantial modifications to the method. For the parametric case, a standard 3-parameter location-scale t -distribution is fitted to the rescaled intraday returns using maximum likelihood. The mathematical details of this approach will not be discussed here, since all of the techniques involved are standard and have been discussed in detail elsewhere.

In the nonparametric case, the standard kernel estimator for a univariate density

function will be employed here (originally due to Parzen, 1962, and sometimes referred to as the Parzen-Rosenblatt kernel estimator). Consistent with previous notation, the observed series of T intraday returns are denoted by $\{Y_{I,t}\}_{t=1}^T$ and the rescaled intraday returns are denoted by $\{Y_{R,t}\}_{t=1}^T$, where $Y_{R,t} = c^{\hat{H}} Y_{I,t}$ and $\hat{H}$ is an estimate of the self-affinity index. The kernel density estimator for the density of the rescaled intraday returns $f(y_R)$ is then given by:

$$\hat{f}(y_R) = \frac{1}{hT} \sum_{t=1}^T k\left(\frac{Y_{R,t} - y_R}{h}\right) \quad (2.3.1)$$

where $k(\cdot)$ is a nonnegative and bounded kernel function and the parameter h is a bandwidth or smoothing parameter. Following the discussion of Section (2.2), for a unifractal process these rescaled intraday returns should be equal in distribution to the daily returns observed over the same time period. Therefore, the estimated density of the rescaled intraday returns, $\hat{f}(y_R)$, provides an estimate of the daily return density $f(y_D)$.

In addition to avoiding the need to select a particular functional form for the density of returns, the use of nonparametric kernel-based estimation methods has the added potential benefit of introducing a certain degree of smoothing into the estimation of the density function. Given that high-frequency financial data often contain a certain amount of noise due to market microstructure effects and other factors, an estimation method that automatically smooths out some of the most extreme or erroneous observations may be advantageous. Indeed, the benefits of smoothing when dealing with intraday financial data have already been demonstrated in the realised volatility literature, through the use of kernel-based estimators of return volatility (see in particular Hansen and Lunde, 2006).

Standard regularity conditions on the kernel function and bandwidth parameter h

(see for example Li and Racine, 2006) guarantee uniform consistency of the kernel estimator in (2.3.1) in the case of independent data. However, it is well known that financial returns at the daily and intraday frequencies considered here typically exhibit some form of serial dependence. However, the same uniform almost sure rate of convergence for the standard kernel estimator is preserved in the case of weakly dependent data, provided that the structure and strength of serial dependence satisfies certain conditions⁶. In the current context, we require the following assumption:

A.1.2 The intraday return process $Y_{I,t}$ observed over each individual trading day is strictly stationary and α -mixing with α -mixing coefficients satisfying $\alpha(j) = O(j^{-(1+\epsilon)})$, for some $\epsilon > 0$.

It should be noted that assumption A.1.2 only requires the conditions on serial dependence to hold for each individual trading day of intraday data in isolation and not for the whole intraday return process over the complete sample period. This is because, as discussed during the following section, the daily return density for each trading day is estimated just from the rescaled intraday returns observed over that trading day and not from multiple days. Intuitively it seems plausible that dependence between the intraday returns in each period and those at the start of trading becomes increasingly small as we move towards the end of the trading day. This argument could fail if multiple days of intraday data were used for estimation, since the patterns of intraday seasonality present in intraday returns may introduce long-range dependence in the intraday return process, which is not permitted by assumption A.1.2. In addition, estimates of the long-range dependence parameter from the complete sample of intraday data and the autocorrelation functions for longer lag lengths provide no strong evidence for the presence of long-range dependence in the level of the intraday return series at longer horizons. Assumptions

⁶Further details of kernel estimation for dependent data, together with a brief summary of relevant mixing conditions, can be found in Li and Racine (2006).

A.1.1 and A.1.2 lead to the following proposition:

Proposition 1.1 Under assumptions A.1.1 and A.1.2, the kernel estimator for the probability density function of rescaled intraday returns given in equation (2.3.1) is a consistent estimator of the probability density function of daily returns.

The proof of Proposition 1.1 follows from standard consistency arguments for nonparametric methods (see for example Li and Racine, 2006). Finally, with kernel-based estimators there are the additional problems of choosing the kernel weighting function $k(\cdot)$ and optimally selecting the value of the bandwidth parameter h , in equation (2.3.1). These topics will be discussed in more detail during the empirical exercise.

2.4 Forecasting the Density of Daily Returns

The current section explores how one-step-ahead out-of-sample density forecasts for daily returns can be produced from estimated daily return densities obtained using the method previously presented in Sections 2.2 and 2.3. In order to produce one-step-ahead out-of-sample density forecasts for daily returns from the estimated daily return densities obtained using the previously presented methods, a dynamic structure must be imposed to describe the evolution of the daily return density over time. For both the parametric and nonparametric approaches of the previous section a simple autoregressive structure is employed, however the implementation necessarily differs depending on which class of estimator is used for the density function.

Section 2.4.1 begins by providing a formal description of the forecasting environment assumed throughout the current section and for the empirical exercise in Section 2.5. Sections 2.4.2 and 2.4.3 present the dynamic structures used for density forecasting in the parametric and nonparametric cases, however the nonparametric case is more complex

to implement the and so warrants more discussion. Finally, Section 2.4.4 discusses how the form of these dynamic structures can be estimated in practice in order to produce forecasts for the daily return density.

2.4.1 The Forecasting Environment

It is assumed that a series of intraday returns are observed over a period of T days, together with a corresponding series of daily returns. A standard rolling window scheme with a window size of m -days is employed to produce one-step-ahead density forecasts for daily returns from this intraday data; specifically, at day m , an estimate of the self-affinity index, H , is produced using the intraday data from day 1 up to day m . This estimate is denoted by $\hat{H}_m$ and is then used, together with the value appropriate value of c for the frequency of intraday data employed, to rescale the intraday data for the same m -day period.

The density of the rescaled intraday data for each of the m days is then estimated using the methods presented in Section 2.3.2 to produce a time series of m estimated daily return densities, denoted by $\{\hat{f}_t(y) : 1 \leq t \leq m\}$, with the estimated density for each trading day produced using only the rescaled returns observed during that day. These m estimated daily return densities are used to produce in-sample estimates of the relevant parameters of the chosen dynamic structure and these estimated values are then used to produce an out-of-sample one-step-ahead forecast for the density of daily returns at time $m + 1$. This density forecast is denoted by $\tilde{f}_m(y)$, with the subscript indicating that the forecast is conditional on the information available at time m (but for use at time $m + 1$).

The estimation window is then rolled forward by one day and the above procedure is repeated using the intraday data from day 2 up to day $m + 1$ to produce a one-step-ahead density forecast for use on day $m + 2$, denoted $\tilde{f}_{m+1}(y)$; an updated estimate of the self-affinity index is produced, denoted by $\hat{H}_{m+1}$, which is then used as before to rescale

the intraday data from day 2 up to day $m + 1$. The density of the rescaled data for each day is estimated using the kernel estimator of Section 2.3.2, resulting in a new set of m estimated daily return densities, denoted by $\{\hat{f}_t(y) : 2 \leq t \leq m + 1\}$. This procedure can be repeated over the rest of the sample to produce a sequence of N out-of-sample one-step-ahead density forecasts for daily returns, where $N = T - m$.

Note that the density estimate obtained in each period will differ from those obtained in the previous period, since the updated estimate of the self-affinity index results in a different rescaling factor for the intraday data and therefore a different set of rescaled intraday observations for each of the trading days. Whilst this rolling estimation of the self-affinity index does slightly increase the computational requirements of the method, it permits the scaling properties of the return process to change over time and so results in more flexible scaling behaviour than the time-invariant scaling of a true unifractal process with a constant global value of the self-affinity index. Such a process that exhibits unifractal distributional scaling properties, but allows the parameters characterising these unifractal properties to change over time, is generally referred to as a multifractional process (not to be confused with multifractal). The financial applications of such multifractional processes have attracted increasing interest in recent years, with two examples being the work of Frezza (2012) and Bianchi et al. (2013).

2.4.2 Dynamic Structure for Density Forecasting: Parametric Case

As previously stated, a simple autoregressive dynamic structure is employed in order to produce one-step-ahead density forecasts from the time series of estimated densities obtained using the methods outlined in the previous section. In the case of parametric specifications for the daily return density this is relatively simple, since the dynamic structure describing the evolution of the return density over time can be imposed via the parameters of the chosen distribution; this allows point forecasts for the distributional parameters to be produced that in turn provide a density forecast for daily returns.

For the chosen location-scale t -distribution each of the three distributional parameters (location, scale and degrees of freedom) can be modelled separately using a standard univariate AR(p) model; in-sample estimates of the relevant autoregressive parameters can be obtained from the m estimated daily return densities over the in-sample period and these parameter estimates can then be used to produce one-step-ahead out-of-sample forecasts for the parameters. The only complication with such an approach is that the forecasted parameter values from an autoregressive structure are not guaranteed to be within the permitted parameter space for the distribution without additional constraints being imposed.

2.4.3 Dynamic Structure for Density Forecasting: Nonparametric Case

When the daily return densities are estimated using nonparametric methods such as the kernel estimator of Section 2.3.2, the approach outlined above is clearly inapplicable. Instead, an alternative method is now developed based on imposing a parametric dynamic structure on the evolution of the complete probability density. It is assumed that the entire density at time $t + 1$ depends on several past densities, with this temporal dependence again assumed to follow an autoregressive structure. Whilst this choice is perhaps slightly ad-hoc, autoregressive structures have been shown to work well in the context of quantiles (see for example the Conditional Autoregressive Value-at-risk or CAViaR model of Engle and Manganelli, 2004).

The simplest case would be to assume that each of the lagged densities have a constant coefficient, making the density of daily returns at time $t + 1$ a simple weighted sum of the return densities from t to $t - p + 1$, which can be expressed as:

$$f_{t+1}(y) = \beta_1 f_t(y) + \beta_2 f_{t-1}(y) + \dots + \beta_p f_{t-p+1}(y) + u_{t+1}(y) \quad (2.4.1)$$

The D subscripts indicating daily return densities have been suppressed for notational simplicity and the error terms $u_{t+1}(y)$ are a martingale difference sequence for all values of y in the domain of $f_{t+1}(y)$. The true daily return density in each period, $f_s(y)$, is unknown, but replacing with the corresponding density estimate obtained using the method of Section 2.3, $\hat{f}_s(y)$, gives:

$$\hat{f}_{t+1}(y) = \gamma_1 \hat{f}_t(y) + \gamma_2 \hat{f}_{t-1}(y) + \dots + \gamma_p \hat{f}_{t-p+1}(y) + v_{t+1}(y) \quad (2.4.2)$$

where again, the error terms $v_{t+1}(y)$ are martingale difference sequences for all values of y . In-sample estimates of the autoregressive parameters $\{\gamma_i : 1 \leq i \leq p\}$ can be produced from the time series of estimated daily return densities, with the resulting estimates denoted by $\{\hat{\gamma}_i : 1 \leq i \leq p\}$. When combined with the estimated daily return densities for periods t to $t-p-1$, these in-sample estimates of the autoregressive parameters can then be used to produce a one-step-ahead out-of-sample density forecast for the density of daily returns for day $t+1$ from:

$$\tilde{f}_t(y) = \hat{\gamma}_1 \hat{f}_t(y) + \hat{\gamma}_2 \hat{f}_{t-1}(y) + \dots + \hat{\gamma}_p \hat{f}_{t-p+1}(y) \quad (2.4.3)$$

Note that $\tilde{f}_t(y)$ is used to denote the forecast of the daily return density made conditional on the information available at time t (but for use at time $t+1$) and the tilde is used to distinguish it from the in-sample density estimate at time t . Clearly constraints need to be imposed on the values of the estimated parameters to ensure that the density forecast produced by (2.4.3) is always a valid probability density, but for the autoregressive structure above this is guaranteed simply by constraining the estimated parameters sum to unity and are all non-negative.

Although the dynamic structure of equation (2.4.1) is simple to interpret and im-

plement, one potential limitation is that it does not allow the dependence between the density at time t and that at time $t - s$ to vary across different regions of the density⁷. It could be generalised by replacing the constant coefficients, $\beta_1, \dots, \beta_p$, with functions of y , but ensuring that the resulting density forecast is a valid probability density function is no longer straightforward and so this is left as a possible area for future research.

In fact, the dynamic structure of (2.4.1) is arguably less flexible than that used for the parametric density function approach, in which the three parameters of the chosen location-scale t-distribution are able to evolve independently over time. As a result, it will be interesting to assess during the empirical exercise whether the additional flexibility in dynamic structure allowed by the parametric density specification outweighs the advantages of employing a nonparametric estimator for the daily return density.

2.4.4 Estimating Autoregressive Parameter Values

For the case of parametrically estimated densities the parameter values for the univariate autoregressive models required to produce forecasts can be estimated straightforwardly using standard techniques. For the case of nonparametrically estimated densities the situation is again more complex and alternative techniques must be developed.

The simplest option is to impose some fixed vector of values for the autoregressive parameters $\{\gamma_q : 1 \leq q \leq p\}$ in (2.4.2) over all time periods. Although this approach may seem overly simplistic, it has been shown in the literature on forecast combination that a simple average of forecasts can perform better than a combination chosen to minimise some statistical loss function (see Timmermann, 2006). Clearly the current problem is not identical to that faced in the forecast combination literature, since the current aim is to optimally combine a number of estimated densities from the same model but different time periods, rather than multiple forecasts from different models but a common time

⁷This limitation was not encountered when a similar autoregressive structure was applied by Engle and Manganelli (2004) in the literature on conditional quantiles, because the autoregressive coefficients were estimated separately for each of the individual quantiles, allowing them to vary across the distribution.

period. However, a simple specification with $\gamma_q = 1/p$ for $q = 1 \dots p$ is included in the empirical exercise of the next section to explore whether the same ad-hoc dynamic structure can also perform well in the current context.

Assuming instead that for a given period we wish to identify the autoregressive parameter vector that produces the most accurate one-step-ahead out-of-sample density forecast, then an appropriate loss function needs to be selected to formally define what constitutes the ‘best’ forecast. Given that a probability density is defined across a range of values, conventional point-based measures of accuracy cannot be applied and an alternative loss function appropriate for probability densities must be employed.

As a loss function the current method employs the well-known logarithmic score, which is closely related to the Kullback-Liebler information criterion (KLIC) and is widely employed in the literature on both density forecast comparison and the optimal combination of density forecasts⁸. The application of the logarithmic score for the problem of optimal density forecast combination (see Hall and Mitchell, 2007) is particularly relevant, since the problem of choosing the parameter values for the simple autoregressive structure of equation (2.4.2) is mathematically equivalent to that of identifying the optimal weights for the commonly studied linear combination of density forecasts.

Denoting the one-step-ahead density forecast for daily returns produced at time $t - 1$ for use in period t as before by $\tilde{f}_{t-1}(y_t)$ and the actual daily return observed at time t by y_t^* , the average logarithmic score over the periods $t = 1, \dots, S$ is given by:

$$\frac{1}{S} \sum_{t=0}^{S-1} \ln \tilde{f}_{t-1}(y_t^*) \quad (2.4.4)$$

Given that better forecasting models should on average assign higher probabilities to the outcome that actually occurred, higher values of the average logarithmic score provide

⁸See for example Mitchell and Hall (2005), Amisano and Giacomini (2007), Hall and Mitchell (2007), Bao, Lee, and Saltoğlu (2007) or Kascha and Ravazzolo (2010).

evidence of superior predictive ability. The autoregressive coefficients in equation (2.4.2) can therefore be chosen in order to maximise the average logarithmic score. More formally, the vector of estimated autoregressive parameters $\hat{\gamma}$, is obtained as the solution to:

$$\hat{\gamma} = \arg \max_{\gamma} \frac{1}{S} \sum_{t=0}^{S-1} \ln \tilde{f}_t(y_{t+1}; \gamma) \quad \text{s.t.} \quad \sum_{i=1}^p \gamma_i = 1 \text{ and } \gamma_i \geq 0 \text{ for } i = 1, \dots, p$$

where the constraints ensure that the resulting density forecast is a valid probability density function and the maximisation problem is solved using a numerical optimisation procedure.

For each of the rolling m -day in-sample estimation windows, an estimate of the autoregressive parameter vector, γ , is obtained as the solution to the above optimisation problem for the $m - p$ in-sample forecasts in the current m -day window. The resulting in-sample estimate of the parameter vector, $\hat{\gamma}$, is then used to produce an out-of-sample one-step-ahead density forecast for daily returns, for use in the following period. Each time the m -day in-sample window is rolled forwards by one day, the estimate of the autoregressive parameter vector is updated using the same optimisation procedure and this new vector is then used to produce an out-of-sample forecast for the following period; this is repeated until one-step-ahead density forecasts have been obtained for all days in the chosen out-of-sample period.

The constraints above on the parameter vector γ are imposed only to guarantee that the resulting forecast is a valid density function, but for autoregressive processes in a standard time series context, constraints must be imposed on the autoregressive parameters in order to guarantee stationarity and ergodicity of the process. However, whilst the structure imposed on the evolution of the density function by equation (2.4.2) is

justified from an intuitive perspective using the concept of an autoregressive structure, as stated above the problem is treated mathematically as one of optimal density forecast combination (see again Hall and Mitchell, 2007) and not as a true autoregressive process. The numerical optimisation method used to select the parameter values is not dependent on the assumption of stationarity and a rolling window estimation environment is employed with the parameters updated in each period and used to produce one-step-ahead (rather than multi-step) forecasts. These factors combined imply that stationarity is not a critical assumption for either the estimation or forecasting stages of the proposed methodology and so additional parameter constraints are not required.

Finally, the method for density forecast comparison employed in the empirical section uses an alternative scoring rule known as the continuous ranked probability score (CRPS) to measure the relative accuracy of competing density forecasts. An equivalent numerical optimisation procedure could be employed to select the values of the autoregressive parameter values based on the CRPS instead of the logarithmic score, however this alternative is much more computationally demanding and does not appear to provide substantial gains in predictive ability for the current dataset.

2.5 Empirical Application

The current section applies the new semiparametric density forecasting framework to both foreign exchange and equity data in order to compare the performance of the resulting density forecasts with those of existing methods. Section 2.5.1 describes the dataset employed for the empirical analysis and Section 2.5.2 details the benchmark density forecasting model used for comparison. Section 2.5.3 discusses the statistical test employed to compare the relative density forecasting performance of the models and finally Section 2.5.4 presents the empirical results.

2.5.1 Data

The data used were obtained from Olsen Associates and consist of intraday 5-minute observations from 3rd January 2007 until 31st December 2010 on the Euro (EUR) and Japanese Yen (JPY) exchange rates against the US Dollar (USD) and the levels of the S&P500 and NASDAQ-100 equity indexes. The choice of 5-minute data was guided by the desire to exploit as much of the potentially valuable intraday information as possible, whilst avoiding the distortions caused by market microstructure effects typically encountered at very short sampling intervals⁹.

The raw price or level data contains all 5-minute intervals in the sample period and so weekends and other non-trading days need to be removed. For the S&P500 and NASDAQ-100 data this is a relatively straightforward task, since these markets have well-defined trading hours with no trading taking place over weekends or on holidays (such as Christmas day and Thanksgiving). The list of non-weekend closures for the S&P500 and NASDAQ-100 was constructed from the historical list of holidays available on the NYSE website. Throughout the sample there were also 9 days for which the market was open, but trading took place for reduced hours (such as the day after Thanksgiving); the analysis was performed with these partial trading days both removed and included, but the choice did not have any significant effect on the results.

For the two exchange rate series the situation is more complex, because although trading takes place 24 hours a day and 7 days a week, over weekends and certain holidays trading slows substantially. Following Andersen et al. (2001), the end of each 24-hour trading day was taken to be 21:00 GMT and the 48-hour weekend periods between 21:05GMT on each Friday and 21:00 on each Sunday were removed from the raw 5-minute series. For most of the NYSE holidays during the sample period both the EUR/USD and JPY/USD markets were open for normal trading hours; only for Christmas Day and New

⁹This problem is also encountered in the literature on realised volatility, where the 5-minute sampling interval has generally been found to be a good compromise between these two factors (see for example Andersen et al., 2001).

Year's Day was trading noticeably slower than normal and so only these holidays were omitted from the exchange rate series. The analysis was also performed with a larger and more comprehensive list of holidays removed from the EUR/USD and JPY/USD series, but as with the partial trading days for the equity index data, this did not significantly influence the results.

This process leaves a sample size of 1008 and 1037 trading days for the equity index and exchange rate series respectively. Continuous 5-minute returns were then constructed from the first difference of the log-price series for each asset, with the first 5-minute return for each day calculated between the closing price in the previous trading day and the opening price in the current day (thus including any overnight or weekend effects). The GARCH models used for density forecast comparison and the statistical method used for forecast comparison both require a daily return series and so daily returns were also constructed for each asset from the last 5-minute price observed in each trading day.

As previously discussed in Section 2.2, the proposed density forecasting method is only strictly valid when the distributional scaling behaviour of the return process is consistent with that of a unifractal process, rather than the more general class of multifractal processes. In practice the method should still be applicable even if this assumption is not satisfied exactly, provided that the distributional scaling behaviour of a unifractal process still provides a good approximation of the true scaling behaviour of the process. Firstly, as previously stated in Section 2.2.2, for the current application it is only necessary that the distributional scaling properties are satisfied over the range of sampling intervals of interest (from 5-minutes up to 1-day) and not for any range of timescales as for a true unifractal process.

Secondly, from Section 2.4.1, the fact that the self-affinity index is permitted to vary over time implies that it is sufficient for the process to be locally unifractal within each of the rolling estimation windows, even if it may have more complex scaling properties

when viewed globally. Chapter 4 of this thesis develops a formal statistical test for distinguishing whether a given sample of data is most consistent with a unifractal or multifractal process that can be applied globally or locally.

Previous work on unifractal and multifractal processes in finance has relied instead on an informal graphical method based on the scaling function, $\tau(q)$, from equation (2.2.5) in order to distinguish between unifractal and multifractal scaling. This graphical testing approach may not be reliable for the smaller samples required to study local scaling properties, however substantial deviations from globally unifractal scaling visible using the graphical approach may still provide some indication as to the validity of the unifractal forecasting method for a given series. Therefore, an initial check of the global scaling properties will be performed here using this informal graphical testing method, with the local scaling properties investigated later in Chapter 4 using the more formal testing methodology developed there.

As previously discussed in Section 2.2.1, the scaling function is strictly concave for a multifractal process and linear for a unifractal process with equation $\tau(q) = Hq - 1$, where H is the self-affinity index. The solid lines in Figure 2.1 are the estimates of $\tau(q)$ for each of the series under the assumption of multifractality, obtained from the complete sample of 5-minute data using the standard partition or structure function approach (see Section 3.3.1 in Chapter 3 or Kantelhardt, 2009 for details). The dashed lines in each sub-plot are the linear scaling functions obtained under the assumption of unifractality, with the self-affinity index (and therefore the slope of $\tau(q)$) estimated using the DMA estimator of Section 2.3.1. These functions are plotted over the domain $0 \leq q \leq 5$, which is a common choice in empirical studies of scaling behaviour in asset returns.

From Figure 2.1 it can be seen that although the estimated scaling functions are strictly concave for all series, suggesting multifractal rather than unifractal distributional scaling, the degree of nonlinearity varies for the different series; it appears to be lowest for the EUR/USD data and highest for the JPY/USD data, with the two equity index

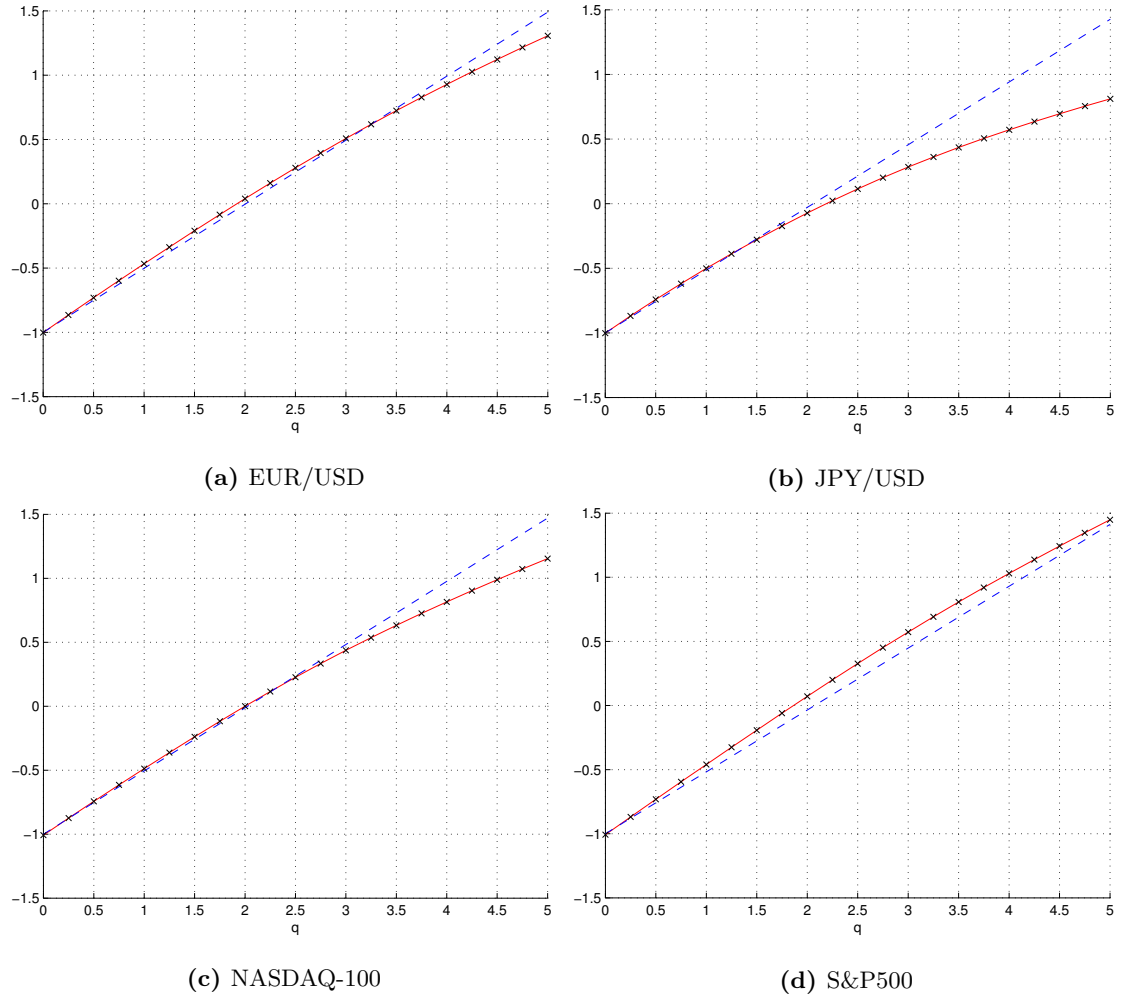


Figure 2.1: Estimated unifractal and multifractal scaling functions. Solid lines correspond to estimated scaling functions for the multifractal case (obtained using the partition function estimator) and dashed lines correspond to the unifractal estimates (obtained using the DMA estimator).

series lying somewhere in between. The differences observed between the estimated scaling functions for the pair of exchange rates and the pair of equity indexes suggest that the type of financial asset under consideration may not in isolation provide any strong a priori information regarding the likely distributional scaling properties of the return process. Whether these observed deviations from unifractal scaling behaviour are sufficiently small for the proposed method to perform well is an empirical issue, which will be considered later in the current section during the density forecasting exercise.

2.5.2 Benchmark Density Forecasting Models

For the empirical exercise it is necessary to have one or more existing density forecasting methods to compare the performance of the new unifractal method against and whilst there are many possibilities, two simple examples have been used initially as benchmarks.

In order to provide a comparison with existing density forecasting methods employing intraday data, the first benchmark method is the autoregressive realised volatility (AR-RV) model of Andersen et al. (2003), which fits a univariate autoregressive model to the time series of (logarithmic, demeaned) daily realised volatility measures. Density forecasts for daily returns can then be produced by combining these point forecasts of volatility with the empirical observation that daily returns are approximately normally distributed if standardised by their corresponding (time-varying) realised volatilities for each day and their constant sample mean. Following Andersen et al. (2003), a 5th order AR-RV(5) model was used initially for the empirical exercise of Section 2.5.4 and this choice was also found to produce the best average density forecasting performance for the dataset employed here. Further details of the AR-RV method can be found in Andersen et al. (2003).

To provide a comparison with methods based on daily data, the second benchmark is a standard GARCH model, with the exponential GARCH and GJR specifications also considered for the equity data to allow for possible leverage effects. A range of

ARMA(p, q) models were tested for the mean equation and for the error distribution the normal, generalised error and Student's t distributions were all tested. Given that the current objective is to forecast the daily return density and not simply the conditional variance, the specifications of the GARCH models were chosen in order to maximise density forecasting performance for the current dataset. In all cases this resulted in GARCH(1,1) volatility equations, AR(1) mean equations and t -distributed errors.

For both benchmark density forecasting methods, the same rolling window estimation scheme as described in Section 2.4.1 was employed for producing density forecasts: the parameters of the models are estimated using an m -day rolling window of data (daily data in the case of the GARCH model and 5-minute intraday data for the case of the AR-RV model) and these parameter estimates are then used to produce one-step-ahead point forecasts for the relevant moments of daily returns. This was then combined with the relevant parametric form assumed for the return distribution to produce one-step-ahead out-of-sample density forecasts for daily returns.

2.5.3 Method for Density Forecast Comparison

The method used for out-of-sample density forecast comparison is the test of equal predictive ability proposed by Gneiting and Ranjan (2011). The test assumes that two competing forecasting models are used to produce one-step-ahead out-of-sample density forecasts for the variable of interest, y . Consistent with the previous notation, it is assumed that N density forecasts are produced by each forecasting method and the forecasts produced by the two models at time t (for use at time $t + 1$) are denoted by $\tilde{f}_t(y)$ and $\tilde{g}_t(y)$, respectively.

The loss function employed by the test is the continuous ranked probability score¹⁰ (CRPS), generalised to allow more importance to be placed on forecast accuracy in

¹⁰The earlier weighted likelihood ratio (WLR) test of Amisano and Giacomini (2007) is similar in spirit, but uses the logarithmic score of Section 3.3. However, it has subsequently been demonstrated that the WLR test is not guaranteed to produce valid inference when a weighting function is used.

particular regions of the density via the use of a weighting function. The value of the weighted CRPS for the forecast produced by the first model for use in period $t + 1$, denoted by $S(\tilde{f}_t, y_{t+1})$, is given by:

$$S(\tilde{f}_t, y_{t+1}) = 2 \int_0^1 \left(\mathbb{I}\{y_{t+1} \leq \tilde{F}_t^{-1}(\alpha)\} - \alpha \right) \left(\tilde{F}_t^{-1}(\alpha) - y_{t+1} \right) w(\alpha) d\alpha \quad (2.5.1)$$

where $\tilde{F}_t(y)$ is the CDF forecast produced at time t obtained from the PDF forecast for the same period $\tilde{f}_t(y)$, $\mathbb{I}\{\cdot\}$ is an indicator function and $w(\alpha)$ is a weighting function; the authors suggest several possible forms for $w(\alpha)$, which allow more weight to be placed on forecast accuracy different regions of the density, such as the centre or tails.

Whenever a closed form expression for (2.5.1) is unavailable, it can be approximated easily to any degree of accuracy using the method outlined by Gneiting and Ranjan (2011). Consistent with the discussion of Section 2.4.1, it is assumed that the original time series of interest is of length T and an m -day rolling window estimation scheme is employed, with the first forecasts produced in period m (for use in period $m + 1$) and the final forecasts produced at time $T - 1$ (for use at time T), thus giving a total of $N = T - m$ out-of-sample forecasts¹¹. The average value of the weighted CRPS in (2.5.1) can be calculated for each of the two density forecasting models over the N out-of-sample periods (from period $m + 1$ until period T) as:

$$\bar{S}^f = \frac{1}{N} \sum_{t=m}^{T-1} S(\tilde{f}_t, y_{t+1}) \quad \text{and} \quad \bar{S}^g = \frac{1}{N} \sum_{t=m}^{T-1} S(\tilde{g}_t, y_{t+1}) \quad (2.5.2)$$

A formal test can then be based on the following test statistic:

¹¹Note that this relationship implies that N/T converges to a non-zero constant, which is required for the sample averages in equation (2.5.2) to make sense.

$$t = \frac{\bar{S}^f - \bar{S}^g}{\hat{\sigma}_n / \sqrt{N}} \quad (2.5.3)$$

where $\hat{\sigma}_n^2$ is a standard heteroskedasticity and autocorrelation consistent estimator for the asymptotic variance of $\sqrt{N}(\bar{S}^f - \bar{S}^g)$.

Under the null hypothesis that the two density forecasting models have equal predictive ability, the test statistic in (2.5.3) is asymptotically normally distributed, with the null rejected at the $\alpha\%$ significance level if $|t| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. Given that lower values of the CRPS correspond to better forecasts, in the case of rejection the forecasting model f should be chosen when the sample value of the test statistic is positive and model g should be chosen when it is negative.

2.5.4 Empirical Results

For the empirical density forecasting exercise two different lengths of rolling in-sample window (values of m in the previous notation of Section 2.4.1) were used for parameter estimation: the first is a relatively typical choice of 250 working days and the second is a much shorter period of 50 working days. In principle, the new semiparametric unifractal forecasting method (and the existing realised volatility based method) could perform better for shorter estimation windows than methods using daily data if it can effectively exploit the additional relevant information contained in the intraday returns. This short 50-day in-sample period has been included in the analysis to establish whether this is the case in practice. In all cases, the density forecasts are compared over the same out-of-sample period to ensure that results are comparable for the different in-sample periods. A 750 working day evaluation period is used, from the 250th until the 1000th

working day in the sample for each series¹².

Throughout the following the new unifractal density forecasting approach is referred to as the autoregressive unifractal (AR-UF) model, with the parametric and nonparametric density variants denoted by AR-UFP(p) and AR-UFNP(p) respectively. As with the benchmark AR-RV model, the optimal order of the autoregressive structures for the AR-UF methods was also typically found to be 5. For the AR-UFNP variant employing the nonparametric kernel density estimator, the choice of kernel weighting function $k(\cdot)$ in equation (2.3.2) had minimal effect and so the standard normal kernel was employed. The simple normal reference rule-of-thumb was used to select the bandwidth parameter; more complex plug-in and cross-validation approaches were also tested, but the former did not produce significant improvements in density forecasting performance and the high computational requirements of the latter made it impractical in the current rolling estimation context.

Finally, the values for the minimum and maximum window sizes for the DMA estimator of the self-affinity index that were found to be approximately optimal in this context were $n_{min} = 5$ for all series and $n_{max} = 100$ and 300 for the equity and exchange rate series respectively. These values of n_{max} coincide approximately with the number of 5-minute intraday returns observed during a single trading day for each series, suggesting that it is optimal to estimate the scaling behaviour of the process just over the range of sampling intervals that are of direct interest for the current application.

Tables 2.1 and 2.2 contain sample values of the CRPS-based test statistic of Section 2.5.3 for equal predictive ability between the new unifractal semiparametric method and the two benchmark density forecasting models. In addition to the simple unweighted version of the test, several of the weighting functions suggested by Gneiting and Ranjan (2011) have been employed here to place more weight on forecast accuracy in the centre, left and right tails respectively (further details of these functions can be found in the

¹²Because of the difference in trading days, the start and end dates of this period are 19th Dec 2007 - 11th Nov 2010 for the two exchange rate series and 31st Dec 2007 - 21st Dec 2010 for the equity indexes.

original paper).

Table 2.1: Density forecast comparison against AR-RV(5) benchmark

Weighting function	None	Centre	Left tail	Right tail
EUR/USD: 250-day in-sample period				
AR-NPUF(5) - estimated AR parameters	-1.291	-1.531	-1.221	-1.020
AR-NPUF(5) - fixed AR parameters	-1.753*	-1.816*	-1.485	-1.474
AR-UFP(5)	-1.337	-1.107	-1.051	-1.487
EUR/USD: 50-day in-sample period				
AR-UFNP(5) - estimated AR parameters	-0.452	-0.650	-0.897	0.061
AR-UFNP(5) - fixed AR parameters	-1.058	-0.987	-1.473	-0.498
AR-UFP(5)	-0.738	-0.587	-1.185	-0.165
JPY/USD: 250-day in-sample period				
AR-UFNP(5) - estimated AR parameters	1.690*	1.673*	1.009	1.987**
AR-UFNP(5) - fixed AR parameters	1.208	1.021	0.167	1.891*
AR-UFP(5)	0.175	0.311	0.156	0.155
JPY/USD: 50-day in-sample period				
AR-UFNP(5) - estimated AR parameters	0.971	1.016	0.522	1.181
AR-UFNP(5) - fixed AR parameters	0.540	0.530	0.506	0.422
AR-UFP(5)	-0.068	0.051	-0.080	-0.037
NASDAQ100: 250-day in-sample period				
AR-UFNP(5) - estimated AR parameters	-0.833	0.446	-0.523	-0.817
AR-UFNP(5) - fixed AR parameters	-1.996**	-0.693	-2.111**	-1.166
AR-UFP(5)	1.580	1.557	1.473	1.356
NASDAQ100: 50-day in-sample period				
AR-UFNP(5) - estimated AR parameters	0.012	0.100	0.191	-0.171
AR-UFNP(5) - fixed AR parameters	-1.562	-1.486	-1.381	-1.190
AR-UFP(5)	-0.111	0.091	0.527	-0.740
S&P500: 250-day in-sample period				
AR-UFNP(5) - estimated AR parameters	-0.752	0.295	-1.327	0.206
AR-UFNP(5) - fixed AR parameters	-1.937*	-1.056	-2.295**	-0.576
AR-UFP(5)	1.338	1.521	1.452	0.808
S&P500: 50-day in-sample period				
AR-UFNP(5) - estimated AR parameters	-1.000	-0.948	-1.358	-0.472
AR-UFNP(5) - fixed AR parameters	-1.646*	-1.718*	-2.295**	-0.413
AR-UFP(5)	-0.318	0.099	0.483	-1.197

The test statistic is normally distributed under the null of equal predictive ability and the test statistic is constructed such that significant negative (positive) values imply the new unifractal (AR-RV benchmark) method provides superior density forecasting performance. Significance at the 10%, 5% and 1% levels is indicated by one, two or three asterisks, respectively.

Table 2.1 presents the results from comparing the density forecast from the AR-UF

models against those from the realised volatility based AR-RV benchmark model. It can be seen that in the majority of cases the null of equal predictive ability cannot be rejected, implying that the new AR-UF approach equals the performance of the existing AR-RV method. Comparing the results across the columns of Table 2.1, there do not appear to be any completely consistent patterns in relative forecasting performance across the regions of the density, although arguably the performance of the AR-UF models against the AR-RV benchmark is somewhat stronger in the left tail of the density. This suggests that the method should perform well in risk management applications, such as the calculation of Value at Risk or expected shortfall.

The relative performance of the AR-UF models appear to be stronger than the AR-RV method for the EUR/USD and S&P500 data, as indicated by the larger proportion of negative sample values for the test statistic and for these two series the method can sometimes provide a statistically significant improvement in density forecasting performance. The situation is however reversed for the JPY/USD data where the majority of the sample values are positive, though almost always too small to be statistically significant.

Table 2.2 contains an equivalent set of results for the case of the GARCH(1,1) benchmark density forecasting model. Perhaps unsurprisingly the relative performance of the AR-UF models improves substantially when switching to this alternative benchmark that utilises only daily data. Across all assets the majority of sample values are now negative and the number of cases in which the AR-UF models provide a statistically significant improvement in density forecasting performance increases. Furthermore, the GARCH benchmark model is never able to provide a statistically significant improvement in predictive ability over the AR-UF models.

Again, for the GARCH benchmark the performance of the AR-UF models is generally much stronger for the EUR/USD data than for the other assets, with the sample values of the test statistic often significant at even the 1% level. This observed variation

Table 2.2: Density forecast comparison against GARCH(1,1) benchmark

Weighting function	None	Centre	Left tail	Right tail
EUR/USD: 250-day in-sample period				
AR-UFNP(5) - estimated AR parameters	-2.686***	-2.532**	-3.321***	-0.752
AR-UFNP(5) - fixed AR parameters	-2.948***	-2.708***	-3.400***	-0.999
AR-UFP(5)	-2.540**	-2.259**	-3.578***	-0.675
EUR/USD: 50-day in-sample period				
AR-UFNP(5) - estimated AR parameters	-2.105**	-2.198**	-2.785***	-1.217
AR-UFNP(5) - fixed AR parameters	-2.545**	-2.430**	-3.220***	-1.575
AR-UFP(5)	-2.370**	-2.207**	-2.962**	-1.432
JPY/USD: 250-day in-sample period				
AR-UFNP(5) - estimated AR parameters	0.071	0.696	0.663	-0.484
AR-UFNP(5) - fixed AR parameters	-0.190	0.229	-0.043	-0.252
AR-UFP(5)	-1.082	-0.373	-0.062	-1.709*
JPY/USD: 50-day in-sample period				
AR-UFNP(5) - estimated AR parameters	-1.290	-0.861	-1.010	-1.254
AR-UFNP(5) - fixed AR parameters	-1.636	-1.219	-1.033	-1.800*
AR-UFP(5)	-1.994**	-1.610	-1.445	-2.174**
NASDAQ100: 250-day in-sample period				
AR-UFNP(5) - estimated AR parameters	-0.393	-0.393	-1.084	0.296
AR-UFNP(5) - fixed AR parameters	-1.143	-1.143	-1.882*	-0.198
AR-UFP(5)	1.012	0.530	0.637	0.050
NASDAQ100: 50-day in-sample period				
AR-UFNP(5) - estimated AR parameters	-2.095**	-2.205**	-2.252**	-1.721*
AR-UFNP(5) - fixed AR parameters	-2.789***	-2.714***	-3.163***	-2.045**
AR-UFP(5)	-1.991**	-1.874*	-1.131	-2.272**
S&P500: 250-day in-sample period				
AR-UFNP(5) - estimated AR parameters	-0.401	-0.045	-0.715	-0.022
AR-UFNP(5) - fixed AR parameters	-1.094	-0.727	-1.910*	-0.115
AR-UFP(5)	1.012	0.530	0.637	1.035
S&P500: 50-day in-sample period				
AR-UFNP(5) - estimated AR parameters	-1.329	-1.355	-1.367	-1.090
AR-UFNP(5) - fixed AR parameters	-2.495**	-2.314**	-2.602**	-1.982**
AR-UFP(5)	-1.584	-1.504	-1.238	-1.505

The test statistic is normally distributed under the null of equal predictive ability and the test statistic is constructed such that significant negative (positive) values imply the new unifractal (GARCH benchmark) method provides superior density forecasting performance. Significance at the 10%, 5% and 1% levels is indicated by one, two or three asterisks, respectively.

in forecasting performance for the unifractal method across the four assets could be attributable to the differences in their distributional scaling properties previously noted

during Section 2.5.1; the deviation from unifractal distributional scaling appears to be smallest for the EUR/USD data and largest for the JPY/USD data, with the equity indexes between these two extremes. This ordering suggests that the predictive ability of the method declines as the distributional scaling properties of the return process move further from that of a unifractal process.

An additional factor that may contribute to this variation in forecasting performance is the difference in the number of intraday observations for each trading day, with the exchange rate series containing over three times more 5-minute observations than the equity index series. Even if the distributional scaling properties of two return series were identical, a greater number of intraday observations should allow more accurate estimates of the return density to be produced and ultimately result in better density forecasts. Given the relatively poor performance of the unifractal method for the JPY/USD series, the issue of sample size does appear to be of secondary importance, however the density forecasting performance for the S&P500 and NASDAQ-100 series might be closer to that for the EUR/USD data if a larger number of intraday observations were available for estimation.

Comparing the performance of the different variants of the AR-UF method it is clear that the simpler fixed parameter variant of the AR-UFNP model performs better than that in which the parameters are chosen to maximise the logarithmic score; although the differences are sometimes small, this result is consistent across the results in Table 2.1. It may seem surprising that the more restrictive variant of the AR-UFNP model can consistently outperform the more flexible specification, but several points should be considered.

Firstly, as previously noted, empirical studies in the literature on density forecast combination often find that a simple average of forecasts performs better than an 'optimal' combination chosen to minimise some statistical loss function. The problem of density forecast combination is closely related to the problem considered here and so

similar results may also hold. The strong performance of a fixed parameter vector placing equal weight on each lagged density is however harder to justify in the current time series context, where it might be expected that the more recent densities would have larger weights, as is typically the case when fitting standard autoregressive models in a time series context. However, in the current application the dynamic structure is being imposed on complete densities and not single observations, so it is not necessarily true that the same patterns should hold here.

Another possible explanation for this is the problem of estimation error; in principle a time-varying autoregressive parameter vector may be able to provide better performance than a simple fixed parameter vector, but if in practice it is not possible to accurately estimate the optimal parameter values then the simple equally weighted version may perform as well or better on average over the out-of-sample period. This issue of estimation error is one of the explanations given in the forecast combination literature for the strong performance of simple forecast averages (see Timmermann, 2006) and the same argument may apply here.

Finally, the loss function used to estimate the autoregressive parameter values is based on the logarithmic score rather than the CRPS employed for forecast comparison and there is no a priori reason to expect the optimal forecasts in a logarithmic score sense to coincide with those in a CRPS sense. It would in principle be possible to estimate the autoregressive parameters using the same CRPS loss function, but there are practical difficulties with this approach. For nonparametric densities a discretised approximation of the CRPS in equation (2.5.1) must be employed using a vector of values in the domain of the density function. This dramatically increases the computational requirements of the parameter estimation algorithm compared to the previous logarithmic score loss function, with the demands of the optimisation problem increasing with the length of the vector of values. For shorter vectors that remain computationally feasible, forecasts from this CRPS-based method actually perform worse than the existing method, despite

being substantially slower to calculate.

The forecasting performance of the final AR-UFP specification that employs a parametric specification for the daily return density is more variable. For the shorter 50-day in-sample period the differences in predictive ability between the AR-UFP and AR-UFNP specifications are typically small, but for the longer 250-day in-sample period the differences are often more significant; for the JPY/USD data the AR-UFP model provides a large increase in predictive ability over the two AR-UFNP specifications, but for the two equity index series the situation is reversed.

Compared to the AR-UFNP specification, the AR-UFP model imposes a more restrictive form for the daily return density for each trading day, but at the same time permits a more flexible dynamic structure by allowing each of the distributional parameters to evolve independently over time, rather than imposing a single autoregressive structure on the complete density function. The differences in relative forecasting performance between these two specifications across different assets suggest that for some return series it is more important to allow for flexibility in the distributional form than in dynamic structure and for others the converse is true.

The CRPS-based test in Tables 2.1 and 2.2 provides a comparison of average predictive ability over the complete out-of-sample period. However, it is also possible to calculate the CRPS differential for each of the days in the out-of-sample period individually to examine whether the relative forecasting performance of the two methods varies systematically over time. This period-by-period CRPS differential for the AR-RV benchmark is plotted in Figure 2.2 over the complete 750-day out-of-sample period for the EUR/USD and S&P500 data, together with the daily realised volatility used as a proxy for the latent daily return variance¹³. The period-by-period CRPS differentials of Figure 2.2 are constructed using the unweighted version of the CRPS and the longer 250-day rolling estimation window. As with the values of the test statistics reported in

¹³Equivalent figures for the GARCH benchmark have been omitted in order to conserve space, but display similar patterns to those for the AR-RV benchmark.

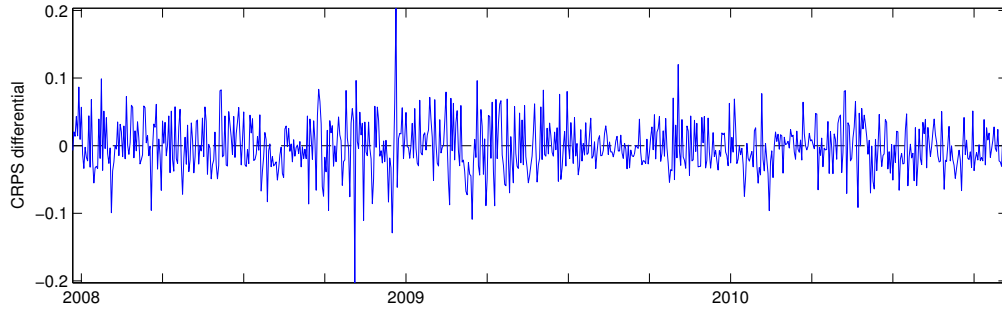
Tables 2.1 and 2.2 above, a negative value of the period-by-period CRPS differential for a given day implies that the unifractal method had superior predictive ability for that day, with larger negative values (in absolute terms) implying a larger improvement over the benchmark AR-RV method.

From Figures 2.2a and 2.2c, it is clear that the predictive ability of the new unifractal forecasting method relative to the AR-RV benchmark varies over the length of the out-of-sample period. Perhaps most notably, for the S&P500 data the relative performance of the unifractal method appears to be stronger during the more volatile period in late 2008 and early 2009, corresponding to the most severe part of the recent financial crisis. This graphical observation is also supported numerically, with the average value of the CRPS differential for the 12 month period from the start of Q3 2008 until the start of Q3 2009 being -0.0137, compared to an average of -0.0033 for the 750 day out-of-sample period as a whole.

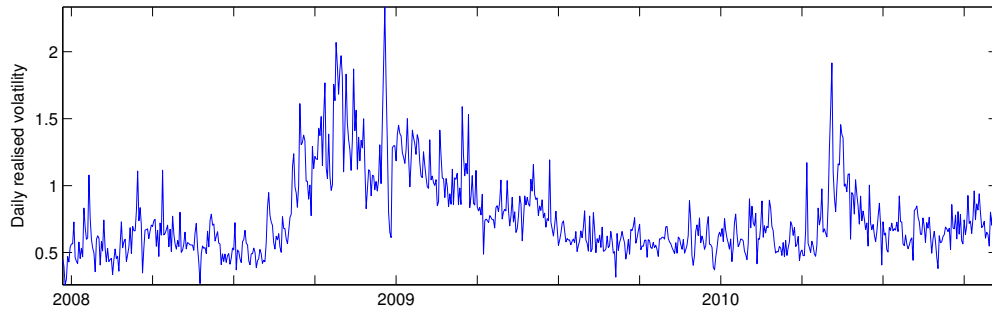
2.6 Conclusion

The current chapter has presented a new method for producing semiparametric density forecasts for daily financial returns using high-frequency intraday data. Through a new application of results from the theory of unifractal processes the intraday returns are appropriately rescaled and the density of daily returns for each trading day is estimated directly from these rescaled high-frequency observations, allowing for the use of both parametric and non-parametric estimators for the daily return density.

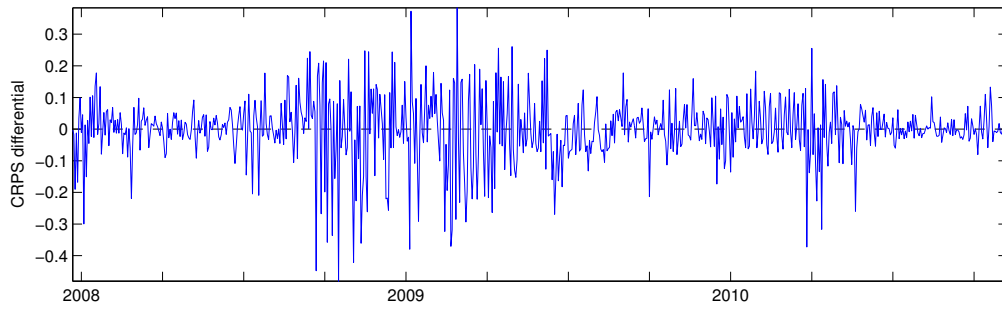
The key assumption required for the proposed method is that the returns exhibit distributional scaling consistent with a unifractal process. It is important to note however that for the proposed method to be applicable, the assumption of unifractal scaling does not need to hold for all timescales and across any sub-period of the data as it does in the classical definition. Instead, it is sufficient for unifractal scaling to be present locally



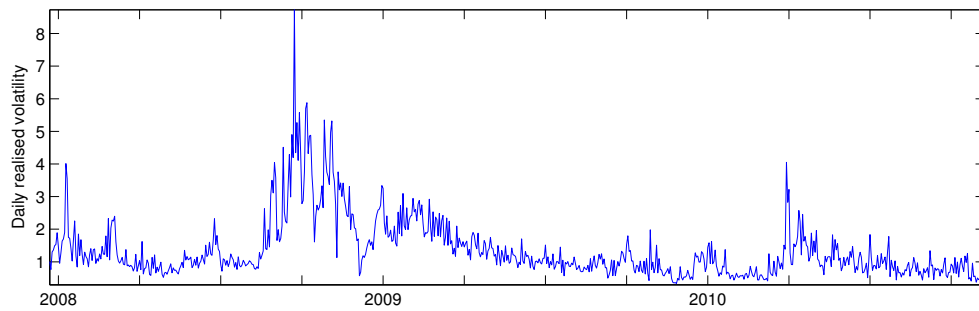
(a) Daily CRPS differential between unifractal and AR-RV density forecasts - EUR/USD



(b) Daily realised volatility - EUR/USD



(c) Daily CRPS differential between unifractal and AR-RV density forecasts - S&P500



(d) Daily realised volatility - S&P500

Figure 2.2: Period-by-period CRPS differential and daily realised volatility.

within each trading day and it may exist only over the range of sampling intervals that are of interest in the intended application of the method (5-minutes to daily in the case of the empirical exercise presented here).

In contrast to previous methods using realised volatility measures to estimate and forecast daily return densities, the approach presented here utilises information about both the sign and magnitude of the intraday returns and allows this intraday information to influence aspects of the daily return density beyond the second moment. In addition, the ability to use nonparametric density estimation techniques avoids the potential difficulties encountered in selecting a suitable parametric model for asset returns.

The density forecasting performance of the new unifractal method was compared against existing methods using both intraday and daily data in an empirical application with 5-minute intraday data on major exchange rates and equity indexes. The empirical results demonstrate that the new unifractal method performs well for return series with distributional scaling properties close to the unifractal scaling required by the method. In particular, for the EUR/USD data the method produces a statistically significant improvement in predictive ability over density forecasts from a standard GARCH model and matches the performance of the autoregressive realised volatility model. For time series that exhibit larger deviations from the required unifractal distributional scaling, such as the S&P500 and NASDAQ-100 return series, the density forecasting performance is typically still competitive with existing methods, particularly for shorter in-sample periods where it is able to again provide statistically significant improvements in predictive ability in several cases. In addition, the gains in predictive ability provided by the new unifractal method seem to increase during periods of high return volatility, such as the most severe part of the financial crisis in late 2008 and early 2009.

The most obvious extension to the work of the current chapter would be to develop an analogous forecasting method under the more general assumption of multifractal distributional scaling, which is the focus of the next chapter. Another potentially interesting

direction for future research would be to explore whether the density of daily returns for a given period, conditional on the return observed in the previous period, could be estimated directly from the intraday data in an analogous way by exploiting similar distributional scaling results. Density forecasts for daily returns could then be produced simply by updating the relevant conditioning information in the estimated conditional density function for each day, thus avoiding the need to impose a parametric dynamic specification in order to produce forecasts.

Chapter 3

Density Forecasts of Daily Financial Returns from Intraday Data II: A Multifractal Approach

3.1 Introduction

In response to the theoretical limitations of existing methods for incorporating intraday data into forecasts for daily returns discussed in Section 2.1, the preceding chapter proposed a method for estimating and forecasting the probability density of daily returns from intraday data, based on a new application of distributional scaling laws for the class of unifractal processes. As discussed in detail there, these processes possess a form of scale invariance, such that the distribution of the process at a given timescale is related to that at any other timescale through a distributional scaling law. The form of this distributional scaling can be estimated for a given sample of data and it was demonstrated how these estimates can be used to appropriately rescale the intraday returns such that they are equal in distribution to daily returns; the density of daily returns can then be directly estimated from these rescaled intraday observations.

In contrast to existing methods, information concerning both the magnitude and sign of intraday returns can be incorporated into the estimates of the daily return density. Furthermore, this approach also allows the use of nonparametric density estimation methods, thus removing the need to impose a specific parametric form for the density of daily returns. The empirical application of Chapter 2 suggests that the proposed unifractal density forecasting method produces density forecasts that perform well when the true scaling behaviour of the return processes is sufficiently close to that of a unifractal process, even if it is not exactly unifractal. However, it also appears that the predictive ability of the unifractal approach can be adversely affected by larger deviations from the unifractal distributional scaling behaviour that is required for the method to be theoretically valid.

The current paper therefore proposes an alternative approach for producing density forecasts for daily returns from intraday data, based on distributional scaling laws for the more general class of multifractal processes. Compared to unifractal processes,

multifractal processes allow for a more flexible scaling relationship between return distributions at different sampling frequencies, overcoming a key theoretical limitation of the previous method. However, whilst the multifractal approach of the current paper permits more flexible distributional scaling behaviour than the earlier unifractal approach, the implementation of the method is more restrictive in some respects, most notably requiring a parametric form to be selected for the daily return distribution. Nonetheless, the proposed method still allows the intraday data to directly influence properties of the daily return density beyond the second moment. In particular, the approach allows the kurtosis of daily returns to be estimated directly from the intraday data and incorporated into the forecasts of the density of daily returns.

The aim of the current paper is therefore to formalise this alternative multifractal approach and explore whether the additional flexibility it permits in terms of distributional scaling behaviour allows it to produce accurate density forecasts, despite the more restrictive implementation it requires compared to the competing unifractal approach. The density forecasting performance of the proposed multifractal approach is compared to that of benchmark models from the GARCH and realised volatility literature, in addition to the unifractal approach of Chapter 2 in an empirical application using a dataset of 5-minute intraday equity and exchange rate data.

The structure of the paper is as follows: Section 3.2 presents the relevant theory on unifractal and multifractal processes and describes how these results can be applied to link the properties of the return process at different sampling frequencies. Section 3.3 then discusses how these concepts can be applied in practice for the multifractal case to estimate and forecast the moments of daily returns and ultimately forecast the daily return density. Section 3.4 presents the empirical application of the new multifractal approach and finally, Section 3.5 concludes.

3.2 Unifractal & Multifractal Processes

In order to estimate the density of daily returns from intraday data, a method for formally linking the characteristics of return distributions across different sampling frequencies is required. For this purpose the previous chapter relied on theoretical results for the class of unifractal processes, with the current chapter instead considering an equivalent method based on the more general class of multifractal processes.

As previously discussed, both unifractal and multifractal processes exhibit forms of *scale invariance*, such that the behaviour of the process observed at one timescale is, after an appropriate transformation, identical in a statistical sense to that observed at another time scale. Although the theoretical properties of these processes were already discussed to some extent in the previous chapter (again with more detailed treatments available in Mandelbrot, Fisher, and Calvet, 1997, Calvet and Fisher, 2002 or Kantelhardt, 2009), the current section summarises the relevant theory again here for convenience, before exploring how these properties can be applied to relate the distributional properties of the return process at the intraday and daily sampling intervals.

3.2.1 A Review of Unifractal and Multifractal Processes

The distributional scaling behaviour of a unifractal or self-affine process can be defined by a simple expression that links the distribution of the process at different sampling intervals. Formally, unifractal or self-affine processes can be defined in the following way:

Definition 3.2.1. A process is said to be *self-affine* or *unifractal* if for some $H > 0$, all $c \geq 0$ and all $t_1, t_2, \dots, t_k \geq 0$ it obeys the distributional scaling relationship

$$\{X(ct_1), X(ct_2), \dots, X(ct_k)\} \stackrel{d}{=} \{c^H X(t_1), c^H X(t_2), \dots, c^H X(t_k)\} \quad (3.2.1)$$

which can be expressed more compactly as:

$$X(ct) \stackrel{d}{=} c^H [X(t)] \quad (3.2.2)$$

If the increments of the process are stationary, then the distributional scaling law of (3.2.2) also holds at the local level:

$$X(t + c\Delta t) - X(t) \stackrel{d}{=} c^H [X(t + \Delta t) - X(t)] \quad (3.2.3)$$

The parameter H is known as the self-affinity index and can be estimated for a specific time series of data using a variety of methods. Common examples of unifractal processes in finance include the standard Brownian motion, for which $H = 1/2$, and also the more general fractional Brownian motion (and the corresponding increment series, the fractional Gaussian noise), for which H is constant but not constrained to be equal to $1/2$. Equations (3.2.2) and (3.2.3) state that the distribution of the process $X(t)$ and the corresponding increment series are, after an appropriate rescaling, identical when the time scale of the process is changed. In the current context this implies that the distribution of returns over different horizons or sampling intervals, for example 1 hour and 1 day returns, are identical after rescaling by a factor that depends on the characteristics of the particular return process (via H) and the difference between the two sampling intervals (via c).

One can also consider the more general class of multifractal processes, which allow for a more flexible relationship between distributions across different sampling frequencies. In the case of a multifractal process, equations (3.2.2) and (3.2.3) can be generalised to:

$$X(ct) \stackrel{d}{=} c^{H(c)}[X(t)] \quad (3.2.4)$$

and

$$X(t + c\Delta t) - X(t) \stackrel{d}{=} c^{H(c)}[X(t + \Delta t) - X(t)] \quad (3.2.5)$$

where the scaling factor c^H has been replaced by the more general function of c , $c^{H(c)}$, allowing for a more flexible scaling relationship between distributions over different sampling frequencies than in the unifractal case. An alternative characterisation of scaling behaviour is often used, particularly in the case of multifractal processes, for which equation (3.2.4) is perhaps somewhat less intuitive than the unifractal analogue of (3.2.2). It can be shown (see for example Mandelbrot et al., 1997) that a stochastic process $X(t)$ with increments $X(t + \Delta t) - X(t)$ is multifractal if these increments are stationary and satisfy:

$$E[|X(t + \Delta t) - X(t)|^q] = c(q)(\Delta t)^{\tau(q)+1} \quad (3.2.6)$$

where $c(q)$ and $\tau(q)$ are deterministic functions of q . The function $\tau(q)$ in (3.2.6) is referred to as the *scaling function* and describes the scaling behaviour for different moments (i.e. values of q) of the absolute increments of the process $X(t)$ for a given range of sampling intervals, Δt . It can be demonstrated (see Calvet and Fisher, 2002) that for a multifractal process the scaling function is non-linear (though always concave with intercept equal to -1), implying that different moments of the absolute increments scale differently with the sampling interval, Δt , than others. For a unifractal process (3.2.6) also holds, but the scaling function is linear and of the form $\tau(q) = Hq - 1$, where H is the same self-affinity index from equations (3.2.1) and (3.2.2). As with the self-affinity

index, H , for a unifractal process, the scaling function can be estimated for a particular time series using various methods (see Kantelhardt, 2009, for a survey of several common estimators).

3.2.2 Application of the Distributional Scaling Laws

Before explaining how the distributional scaling laws for multifractal processes can be applied it is beneficial to begin with a brief summary of the approach proposed in the previous chapter for the unifractal case, in order to emphasise the differences between the methods and explain why an identical approach cannot be used in the multifractal context.

Assume that a series of T intraday returns are observed for a financial asset over a given period and denote this set of intraday returns and their corresponding probability density function by $\{Y_{I,t} : 1 \leq t \leq T\}$ and $f(y_I)$ respectively, where the I subscript is used to indicate returns at the intraday frequency. Denote the density of daily returns for the same asset over the same time period by $f(y_D)$.

Under the assumption that the return process is unifractal¹, it must satisfy the distributional scaling laws of equation (3.2.2) and (3.2.3), which in the current context imply that:

$$f(y_D) = f(c^H y_I) \quad (3.2.7)$$

From (3.2.7), the density of daily returns is equal to the density of the intraday returns, when these intraday returns have been appropriately rescaled by a factor consisting of two components: c and the self-affinity index, H . From (3.2.1), the value of c is determined solely by the relative lengths of the two sampling intervals and the self-

¹As explained in the preceding chapter, exact unifractal properties are not actually required in practice. Refer to the discussion there for more details

affinity index, H , can be estimated from the intraday data using various estimators².

The estimate of the self-affinity index, denoted by $\hat{H}$, can then be combined with the appropriate value of c to rescale the intraday returns by the factor $c^{\hat{H}}$; the density of these rescaled intraday returns can then be estimated and finally from (3.2.7) the resulting estimate can be viewed as an estimate of the density of daily returns over the same period. The large number of rescaled intraday returns obtained over even short time periods allows a wide range of methods to be used to estimate the daily return density from these observations, including nonparametric methods, such as kernel estimation, as used in the previous chapter.

It is therefore relatively straightforward to estimate the density of daily returns from intraday data when the return process is assumed to be unifractal. In the multifractal case the direct analogue of the distributional scaling rule in (3.2.3) employed above for the unifractal case is given by (3.2.5), with the simple scalar H replaced with the function $H(c)$. If this function could be estimated in an analogous manner to H in the unifractal case, then it would be possible to proceed in the same way as before using an estimate of the function $H(c)$ to rescale the intraday returns. Unfortunately there is no existing method for estimating this function and so any application in the multifractal case must be based on an alternative representation of scaling behaviour.

Instead, we will employ the moment scaling property of multifractal processes given in (3.2.6), which has been widely used in empirical studies of multifractal processes in finance as the basis for estimating the scaling function, $\tau(q)$. However, the resulting estimates of the scaling function have only been used to assess whether the distributional scaling properties of the return process are consistent with that of a multifractal process and have not been employed to estimate the moments of the time series process at one sampling interval from data observed at a different timescale.

One possible reason for this is that the moment scaling condition is not immediately

²A detailed survey of common estimators can be found in Kantelhardt (2009)

applicable in the this context without making some additional assumptions; in particular, equation (3.2.6) describes how the non-central moments of the absolute increments of the price process scale with the sampling interval, but what is of more interest for financial returns is the scaling behaviour of the central moments (such as variance, skewness and kurtosis) of the untransformed returns (i.e. the increments of the log price process).

If it can be assumed that the expected value of returns is zero, then the central and non-central moments are equal. Furthermore, for all even values of q in (3.2.6), the moments of the increments and absolute increments are equal. Therefore, under the assumption of a multifractal return process with mean of zero, from equation (3.2.6) the q -th central moment of the return process at sampling interval Δt , denoted by $m(q, \Delta t)$, is given by:

$$m(q, \Delta t) = c(q) \Delta t^{\tau(q)+1} \quad (3.2.8)$$

for all even values of q . The scaling function, $\tau(q)$, and the prefactor, $c(q)$, can both be estimated from a given sample of intraday data using the method presented in Section 3.3.1 below. These values can then be used to produce an estimate of the q -th central moment of returns at any sampling interval for all even numbered values of q . In particular, this allows both the variance and kurtosis of daily returns for a given time period to be estimated from intraday data observed over the same period.

Unlike the unifractal approach of the preceding chapter, this method does not produce a sample of rescaled intraday data from which the daily return density can be estimated. Instead, a parametric distributional form can be assumed that is uniquely determined by the moments estimated from the intraday data, with some possible candidates discussed in Section 3.3.3.

3.3 Estimating and Forecasting the Moments of Daily Returns from Intraday Data

The current section demonstrates how the theoretical results from the previous section can be used in practice to estimate the moments of daily returns from intraday data under the assumption that the return process is multifractal, before proceeding to the problem of forecasting the moments and density of daily returns. Section 3.3.1 begins by describing the chosen method for estimating the scaling function, $\tau(q)$, and the prefactor, $c(q)$, from a given sample of intraday data, which can then provide estimates of the moments of the return process at the daily sampling interval. Section 3.3.2 then moves to a dynamic context and considers how the moment estimates produced in this way can be used to produce out-of-sample forecasts for the daily return moments and finally Section 3.3.3 discusses a possible method for constructing density forecasts for daily returns from these point forecasts for the daily return moments.

3.3.1 Estimation of the Multifractal Scaling Function

Estimating the moments of daily returns from intraday data requires estimates of the scaling function, $\tau(q)$, and the prefactor, $c(q)$, for the relevant values of q . Whilst many methods have been proposed for estimating the scaling function the majority of them do not provide an estimate of the prefactor, since this is typically not of direct interest in most previous studies of multifractal processes in finance or elsewhere, which tend to focus exclusively on the scaling function. However, from the discussion in Section 3.2.2 it is clear that it is required for the current application in order to estimate the moments via equation (3.2.9)

Initially the partition function estimator was employed for estimation; this is one of the simpler estimators for $\tau(q)$, but was selected primarily because it provides a direct estimate of the prefactor, in addition to being one of the most common estimators em-

ployed in the multifractal finance literature (see for example Calvet and Fisher, 2002). Subsequently, more complex estimators for the scaling function were also tested including the multifractal detrended fluctuation analysis method of Kantelhardt et al. (2002) and the multifractal detrended/centred moving average method of Schumann and Kantelhardt (2011), both of which are discussed in detail in Chapter 4. These more complex methods do not however estimate the prefactor and so the estimates of $\tau(q)$ obtained from these alternative estimators were combined with the corresponding partition function method estimate of $c(q)$. Interestingly however, the density forecasting performance of the method with these more complex alternative estimators was found to be worse than when using the simpler partition function approach and so they were not used in the current chapter.

The partition function method is based directly on the multifractal moment scaling condition of (3.2.6), which must be satisfied by any multifractal process. If the process $X(t)$ is observed over the interval $[0, T]$ and this interval is divided into N subintervals of length Δt then the q -th order *partition function* of $X(t)$ is defined as:

$$S_q(T, \Delta t) \equiv \sum_{i=0}^{N-1} |X(i\Delta t + \Delta t) - X(i\Delta t)|^q$$

From the stationarity of the increments of $X(t)$ it follows that:

$$E[S_q(T, \Delta t)] = N \cdot E[|X(i\Delta t + \Delta t) - X(i\Delta t)|^q]$$

Then from the multifractal moment scaling condition of (2.6) and the fact that $N\Delta t = T$:

$$\begin{aligned}
E[S_q(T, \Delta t)] &= Nc(q)(\Delta t)^{\tau(q)+1} \\
\log E[S_q(T, \Delta t)] &= \log c(q) + \log T + \tau(q)\log(\Delta t) \\
\log E[S_q(T, \Delta t)] &= c^*(q) + \tau(q)\log(\Delta t)
\end{aligned} \tag{3.3.1}$$

where the intercept given by $c^*(q) = \log c(q) + \log T$. Therefore, by calculating the value of $S_q(T, \Delta t)$ for a range of sampling intervals, Δt , it is possible to estimate the value of the scaling function, $\tau(q)$, for a given value of q via equation (3.3.1) from the slope of $\log S_q(T, \Delta t)$ plotted against $\log(\Delta t)$. The corresponding value of the prefactor, $c(q)$, can be estimated via the intercept. In practice this process requires a minimum and maximum sampling interval (i.e. value of Δt) to be selected. This is largely an empirical issue, with the optimal choices being dependent on the intended application of the estimated scaling function and also to some extent on the characteristics of the time series in question; as such, this issue will be discussed further during the empirical exercise of Section 4.

Fisher, Calvert & Mandelbrot (1997) used OLS to obtain estimates of the slope and intercept for each partition function and this is the method that has been employed in the literature since³. Typically this process is repeated for a range of values of q , producing estimates of a set of points on the scaling function; an estimate of the complete function $\tau(q)$ can then be obtained by fitting a curve to this set of points. For the current application this only needs to be performed for the values of q corresponding to the moments of interest. Assuming that we wish to estimate both the variance and kurtosis of daily returns, then the second and fourth central moments are required, which can be

³Note that the value of T corresponds to the length of the original time series used for estimation and so is not a function of the interval size, Δt . As such, the presence of $\log T$ within the intercept term introduces no issues for OLS estimation of the parameters.

obtained from the estimated values of $c(q)$ and $\tau(q)$ for $q = 2$ and $q = 4$. In principle higher order moments can also be estimated in the same way, but it has been noted (see Schmitt et al., 1999) that estimates of the scaling function from a finite time series will become less reliable as the value of q increases. Given that these higher moments have less direct interpretation in finance, attention will be restricted to the second and fourth moments.

3.3.2 Forecasting the Moments of Daily Returns

The discussion so far has only considered the estimation of the moments of daily returns from intraday data in a static context, but given the final objective of forecasting, extending this to a dynamic environment is required. This can be achieved by applying the above estimation method to a rolling window of intraday data; by rolling this estimation window forward one day at a time, a time series of estimates for the daily return variance and kurtosis is obtained, with an estimate of each moment for every trading day.

More formally, it is assumed that a series of intraday returns are observed over a period of T days, together with a corresponding series of daily returns. At day m , estimates of the scaling function and prefactor are produced using the first m days of intraday data (from day 1, up to day m) and are then used to estimate the variance and kurtosis of daily returns for day m via equation (3.2.8). The m day window is then rolled forward by one day and the above procedure is repeated using the intraday data from day 2 up to day $m + 1$ to produce estimates of daily return variance and kurtosis for day $m + 2$. By repeating this process over the complete sample, a time series of $M = T - m + 1$ estimates for both the daily return variance and kurtosis are obtained.

Producing moment forecasts from these time series of estimated moments requires some form of dynamic structure to be imposed that describes the evolution of the daily return process over time. The simplest way of achieving this is to impose the dynamic structure directly onto the time series of estimated moments themselves; this is an

approach previously employed in the realised volatility literature, where various time series models have been fitted to daily realised volatility measures obtained from intraday data in order to produce forecasts for daily volatility.

Numerous time series models have been employed for this purpose in the realised volatility literature, with some allowing for relatively complex dynamics, such as the Mixed Data Sampling (MIDAS) and Heterogeneous Autoregressive models used by Clements et al. (2008). Whilst these could also be employed here, two simpler autoregressive specifications will be considered initially that were previously used by Andersen et al. (2003) to model and forecast the realised volatility of three exchange rates.

The first possibility is to assume that the dynamics of the daily return variance and kurtosis can each be described separately by a standard univariate autoregressive (AR) model. Whilst the true values of the daily return variance and kurtosis for day t are not observable, they can be replaced by their corresponding multifractal estimates obtained from the intraday data using the method of the previous subsection. These multifractal moment estimates of the daily return variance and kurtosis are denoted by $\hat{\sigma}^2$ and $\hat{k}$ respectively, with the hats used to emphasise the fact that we are modelling the observable estimated daily return moments and not the true latent moments of the daily return process. The general form of the first model is then given by:

$$\begin{aligned}\log \hat{\sigma}_{t+1}^2 &= \alpha + \sum_{i=0}^{p-1} \phi_i \log \hat{\sigma}_{t-i}^2 + \epsilon_t \\ \log \hat{k}_{t+1} &= \beta + \sum_{j=0}^{q-1} \psi_j \log \hat{k}_{t-j} + \nu_t\end{aligned}\tag{3.3.2}$$

where ϵ_t and ν_t are iid error terms. The specification in (3.3.2) will be referred to as the autoregressive multifractal variance and kurtosis, or AR-MFVK(p, q) model. Note that following Andersen et al. (2003), the logarithmic multifractal moment estimates are

modelled in practice, rather than their levels, for two reasons. Firstly, the logarithmic multifractal moment estimates are much closer to being normally distributed than their levels and so should be easier to model using standard Gaussian time series methods; this is supported in practice when testing for dynamic misspecification using the standard Ljung-Box test for residual serial correlation, where models based on the levels of the estimated moments display more evidence of dynamic misspecification than equivalent models based on the logarithmic moments. Secondly, this guarantees that the resulting moment forecasts obtained are non-negative, as is required for the second and fourth standardised moments.

The parameters in (3.3.2) can then be estimated from the time series of estimated moments and used to produce one-step-ahead forecasts for the daily return variance and kurtosis. Denoting these parameter estimates by $\hat{\alpha}$, $\hat{\beta}$, $\{\hat{\phi}_i : 1 \leq i \leq p\}$ and $\{\hat{\psi}_j : 1 \leq j \leq q\}$, one-step-ahead out-of-sample forecasts for the moments at time $t + 1$ are then given by:

$$\begin{aligned}\log \tilde{\sigma}_{t+1}^2 &= \hat{\alpha} + \sum_{i=0}^{p-1} \hat{\phi}_i \log \hat{\sigma}_{t-i}^2 \\ \log \tilde{k}_{t+1} &= \hat{\beta} + \sum_{j=0}^{q-1} \hat{\psi}_j \log \hat{k}_{t-j}\end{aligned}\tag{3.3.3}$$

where tilde is used to distinguish the one-step-ahead out-of-sample forecasts for the moments, from the in-sample multifractal moment estimates. A slightly more general dynamic structure can be considered that allows for interdependence between the two moments by jointly modelling the daily return variance and kurtosis using a vector autoregression (VAR); this is again similar in spirit to the trivariate VAR specification previously used by Andersen et al. (2003) to jointly model and forecast the realised volatilities of three exchange rates. The general p -th order form of the model, expressed

in terms of the estimated moments, is given by:

$$\begin{pmatrix} \log \hat{\sigma}_{t+1}^2 \\ \log \hat{k}_{t+1} \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \sum_{i=0}^{p-1} \begin{pmatrix} \phi_{11,i} \phi_{12,i} \\ \phi_{21,i} \phi_{22,i} \end{pmatrix} \begin{pmatrix} \log \hat{\sigma}_{t-i}^2 \\ \log \hat{k}_{t-i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \quad (3.3.4)$$

The dynamic specification of (3.3.4) was also tested in the empirical exercise of Section 4, but was found to produce nearly identical density forecasting performance to the simpler pair univariate AR models in equation (3.3.2) and as a result, this alternative VAR specification has been omitted when reporting the empirical results.

Consistent with the previous notation, it is assumed that M multifractal estimates can be produced for the variance and kurtosis from the complete sample of T days of data; for notational simplicity it will be assumed from this point onwards that the first of these moment estimates are produced for period 1⁴.

A standard rolling estimation scheme is used for producing out-of-sample forecasts for the daily return variance and kurtosis, with an n -day in-sample window used to estimate the values of the parameters in (3.3.2) or (3.3.4). In period n the values of the parameters in (3.3.2) or (3.3.4) are estimated using the first n estimated moments of order 2 and 4 (from period 1 to period n) and these parameter estimates are then substituted into (3.3.2) or (3.3.4) to produce one-step-ahead forecasts of variance and kurtosis for use in period $n + 1$. The n -period in-sample window is then rolled forward by one day and the n moment estimates from period 2 up to period $n + 1$ are used to produce moment forecasts for use in period $n + 2$. This process can be repeated to produce one-step-ahead moment forecasts for each day in the chosen out-of-sample period.

Finally, it should be noted that the use of the multifractal moment estimates to approximate the unobserved true daily return moments in the predictive regressions of

⁴Following the previous discussion, this will not automatically be the case unless $m = 1$.

(3.3.2) and (3.3.4) can potentially present difficulties for inference due to the ‘generated regressor problem’ of Pagan (1984). However, the use of a fixed length rolling estimation window means that this should not be a problem in the current context, since the length of the estimation sample does not grow to infinity.

3.3.3 Producing Density Forecasts from Point Forecasts of Moments

Whilst the forecasts of the variance and kurtosis of daily returns could be used directly in many financial applications, the aim of the current paper is to produce forecasts of the complete probability density. In the multifractal case, the obvious way to achieve this is to impose a specific parametric distribution for daily returns that is uniquely characterised by the forecasted moments.

As discussed previously in Section 3.2, one of the limitations of the multifractal approach is the inability to estimate the odd-numbered moments of daily returns, such as skewness, from intraday data. Initially attention will simply be restricted to symmetric distributions, as is common in financial econometrics, but non-zero values could be imposed for any odd-numbered moments based on estimates from daily data or other information. A key advantage the multifractal method possesses over existing methods based on realised volatility measures is that it allows the daily return kurtosis to be estimated directly from intraday data, in addition to the variance. Therefore a symmetric parametric distribution is required for daily returns, which will also allow kurtosis to vary and be determined independently of the variance; this eliminates the normal distribution and the simple 1-parameter version of the t -distribution as suitable options. The generalised error distribution is also unsuitable, since recovering the distributional parameters from the moments requires inversion of the Gamma function⁵.

An obvious choice commonly employed for modelling financial returns is the more general location-scale (or three-parameter) t -distribution, which has the density function:

⁵This is only possible as an approximation and only then for values of the distributional parameters that produce an unsuitable density function for modelling asset returns.

$$f(x : \mu, \lambda, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\lambda\sqrt{\nu\pi}\Gamma(\nu/2)} \left[\frac{\nu + \left(\frac{x-\mu}{\lambda}\right)^2}{\nu} \right]^{-(\nu+1)/2}$$

where μ, λ and ν are the location, scale and degrees of freedom parameters respectively. The distribution has mean equal to μ and skewness equal to zero; the variance and the fourth central moment, denoted by σ^2 and m_4 , are given by:

$$\sigma^2 = \frac{\lambda^2\nu}{\nu-2} \quad \text{for } \nu > 2 \quad \text{and} \quad m_4 = \frac{3\lambda^4\nu^2}{(\nu-4)(\nu-2)} \quad \text{for } \nu > 4$$

Kurtosis, denoted by k , is then equal to:

$$k = \frac{m_4}{\sigma^4} = \frac{3(\nu-2)}{(\nu-4)} \quad \text{for } \nu > 4$$

The distribution automatically satisfies the assumption of symmetry, but the location parameter μ must also be set equal to zero to satisfy the assumption that the mean of returns is zero required by the multifractal approach. The degrees of freedom and scale parameters, ν and λ , can then be obtained from the estimates of daily return variance and kurtosis produced using the intraday data via:

$$\nu = \frac{4k-6}{k-3} \quad \text{for } 3 \leq k \leq 9 \tag{3.3.5}$$

and

$$\lambda = \sqrt{\frac{\sigma^2(\nu-2)}{\nu}} \tag{3.3.6}$$

From equations (3.3.5) and (3.3.6) the location scale t -distribution allows the values of the distributional parameters to be recovered easily from the multifractal estimates of the variance and kurtosis, although kurtosis is required to satisfy $3 \leq k \leq 9$ in order for the distributional parameters to be well-defined. Whilst this restriction is generally satisfied for daily return kurtosis estimates from long samples of intraday data, the lower bound of $k = 3$ can be violated by the multifractal estimates of kurtosis produced from short windows of intraday data. The simplest way to overcome this problem is to impose a lower bound on kurtosis of $k = 3$, so that if the estimated value of kurtosis is strictly less than 3, then it is truncated and set equal to 3, resulting in a normal distribution (or equivalently a location-scale t -distribution with infinite degrees of freedom).

Given that there is no theoretical justification for this restriction and that imposing it discards information whenever the estimated value of kurtosis is less than 3, alternative distributional forms were also explored. The two most notable alternatives considered were the Pearson distribution family and the Gram-Charlier expansion. The former also results in a location-scale t -distribution when kurtosis is greater than 3, but a symmetric 4-parameter beta distribution otherwise; such a distribution has finite support and performs poorly when used to model asset returns. The Gram-Charlier expansion is a semi-parametric method in which the moments, such as skewness and kurtosis, appear directly as parameters. However, the polynomial expansion is not guaranteed to produce a valid density function unless restrictions are imposed on the moments; for a symmetric distribution this requires kurtosis to satisfy $3 \leq k \leq 7$ (see for example Jondeau and Rockinger, 2001), which is even more restrictive than for the truncated location-scale t -distribution approach proposed above. The resulting density forecasts from both of these alternatives perform worse for the current dataset than the location-scale t -distribution, despite the slightly arbitrary restriction on kurtosis that is required for the parameters to be well defined. This is therefore the method used for the remainder of the paper to construct daily return density forecasts from the values of variance and kurtosis, however

there are clearly potential gains in estimation and forecasting performance available if an alternative distributional form can be found that does not require such restrictions.

3.3.4 Two Sources of Multifractality and an Extension to the Method

A fundamental property of multifractal processes is that the scaling behaviour of fluctuations of different sizes is characterised by a range of scaling exponents and cannot be described by a single scaling exponent as in the unifractal case. As discussed in the literature (see for example Kantelhardt et al., 2002 or Kantelhardt, 2009), these differences in scaling behaviour for different sized fluctuations can arise from two possible sources; the first is multifractality due to a broad probability density function for the process (such as a power-law probability density function), whilst the second is caused by small and large fluctuations of the process having different long-range correlations and is therefore a consequence of the temporal structure of the data. Multifractality of the second type will be eliminated if the time series is shuffled randomly, since any temporal dependence present in the original ordered time series will be destroyed. If the multifractality displayed by the original series is purely of the second type then the resulting shuffled series will display non-multifractal distributional scaling behaviour, if it is entirely of the first type then the scaling behaviour of the shuffled series will be unchanged and finally if both types are present in the original data then the shuffled series will still display multifractal scaling, but weaker than that of the original series.

This issue has previously been studied both for simulated multifractal processes (see again Kantelhardt et al., 2002) and also return series for various financial assets (see for example Onali and Goddard, 2009). In the case of financial data, it is generally found that the randomly shuffled returns display different multifractal scaling behaviour than the original ordered return series (as indicated by differences in the shape of the estimated scaling function) implying that at least some of the scaling present in financial data is due to the second source of multifractality. This in turn implies that the multifractal

estimates of the daily return variance and kurtosis obtained from ordered and reshuffled financial data will generally differ; a potentially interesting extension is therefore to explore which of these moment estimates results in forecasts with the greatest predictive ability, by applying the proposed multifractal method to both ordered and randomly shuffled data. If the part of multifractal scaling due to the temporal structure of the data is not relevant for the current application, then randomly shuffling the intraday data before estimating the daily return moments may result in more accurate density forecasts. Conversely, if this component of scaling behaviour is informative for the current application, then eliminating it through reshuffling the data should reduce the forecasting performance of the multifractal method.

One potential problem with this extension to the method is that the partition function estimator of Section 3.3.1 is affected by the ordering of the observations and so each random shuffling of the data will produce a different estimate of the scaling function and therefore different estimates of the daily return moments. The predictive ability of density forecasts produced from these multifractal moment estimates will therefore become stochastic. This has not posed a problem in previous studies, since scaling functions were only estimated in a static context from very long time series of data, making the resulting estimates of the scaling function relatively insensitive to the specific ordering of the observations obtained from shuffling the data. Unfortunately, in the current context where the scaling function is estimated from a short rolling window of data this is no longer the case and estimated scaling functions obtained from successive reshufflings of each window of intraday data exhibit substantial variability.

The solution proposed here is to shuffle the sample of intraday data multiple times, each time producing a new estimate of the daily return variance and kurtosis, before taking an average of these moment estimates that is then used as before to produce density forecasts for daily returns. However, for this approach to work in practice the average moment estimates obtained over the repetitions must have a tendency to con-

verge to a particular value as the number of repetitions increases. Whether this is the case in practice will be investigated for the current dataset during the empirical exercise of Section 3.4.

3.4 Empirical Exercise

The current section compares the density forecasting performance of the new multifractal approach with that of existing methods when applied to both foreign exchange and equity data. Section 3.4.1 describes the dataset employed for the empirical analysis and Section 3.4.2 discusses the alternative density forecasting methods used as benchmarks to compare the multifractal method against. Section 3.4.3 outlines the methods used to formally compare the relative performance of these competing density forecasting models and finally Section 3.4.4 presents the empirical results.

3.4.1 Data

The dataset used for the empirical exercise of the current chapter is identical to that employed in Chapter 1, both in terms of the raw dataset and also the methods used to prepare the data for the empirical exercise. Full details of the methods used and issues encountered during the preparation of the data can be found in Section 2.5.1, but the key aspects are summarised here for convenience.

The data contain intraday 5-minute observations from 3rd January 2007 until 31st December 2010 on the Euro (EUR) and Japanese Yen (JPY) exchange rates against the US Dollar (USD) and the levels of the S&P500 and NASDAQ-100 equity indexes. From these raw price series weekends and other non-trading days were removed; for the S&P500 and NASDAQ-100 data this consisted of removing all weekends and the non-weekend closures in historical list of holidays available on the NYSE website. For the exchange rate series, which have 24-hour trading 7 days a week, periods of slow trading

over weekends and some holidays were removed. The end of each 24-hour trading day was taken to be 21:00 GMT and the 48-hour weekend periods between 21:05GMT on each Friday and 21:00 on each Sunday were removed, together with Christmas Day and New Year's Day, which were the only holidays during which trading was noticeably slower than normal.

This process leaves a sample size of 1008 trading days for the equity index series and 1037 for the exchange rate series. Continuous 5-minute returns were then constructed from the first difference of the log-price series for each asset, with the first 5-minute return for each day calculated between the closing price in the previous trading day and the opening price in the current day (thus including any overnight or weekend effects). A daily return series was also constructed for both assets from the last 5-minute price observed in each trading day. This daily return series is required for estimating the GARCH models used for density forecast comparison and the statistical method used for forecast comparison.

As in the previous chapter, before proceeding it is worth checking that the assumptions of Section 3.2.2 required for the proposed method to be applicable are satisfied for the current dataset. The first requirement is that the distributional scaling properties of the data are consistent with either a multifractal or unifractal process, in order for the moment scaling condition of equation (3.2.6) to hold. As with the unifractal approach of the previous chapter, this restriction on the distributional scaling structure of the process only needs to hold locally within each of the estimation windows because of the dynamic estimation environment employed.

As a result, the same caveats concerning the distinction between global and local scaling properties discussed in Section 2.5.1 are also true here, with sample estimates of the scaling function produced from the complete sample of data not necessarily being informative about the more relevant local scaling properties. Nonetheless the estimated scaling functions for the current dataset that were previously presented in Section 2.5.1

are reproduced again below in Figure 3.1; the solid lines are estimates of $\tau(q)$ for each series, obtained from the complete sample of 5-minute data using the partition function estimator of Section 3.3.1. These estimates are unrestricted in the sense that no a priori assumptions are made on the shape of $\tau(q)$ and thus strict concavity of the resulting estimates can be viewed as evidence in favour of multifractal distributional scaling.

The second assumption that must be satisfied is that mean value of the return process is equal to zero, to guarantee that the central and non-central moments are equal and equation (3.2.8) holds. The validity of the assumption was checked for the current dataset using the standard test statistic for testing that the population mean is equal to some hypothesised value and for none of the 5-minute return series could the null that the population mean is equal to zero be rejected at any conventional significance level.

3.4.2 Benchmark Density Forecasting Models

Three benchmark density forecasting methods are used in the empirical exercise to compare the performance of the proposed multifractal method against. The first two benchmarks are provided by the same GARCH and autoregressive realised volatility (AR-RV) models previously used for the empirical exercise of Chapter 2. A more detailed discussion of these methods can be found in Chapter 2 in Section 2.5.2, but for convenience they are summarised again here. The GARCH benchmark model employs GARCH(1,1) volatility equations, AR(1) mean equations and t -distributed errors for all of the financial assets included in the empirical exercise. As described in Chapter 2, these specifications were chosen from several standard choices in order to maximise the performance of the resulting density forecasts for the current dataset in terms of the CRPS-based test described below. The AR-RV model fits a univariate autoregressive model to the time series of (logarithmic) daily realised volatility measures, with density forecasts for daily returns produced by combining these point forecasts of volatility with the empirical observation that daily returns are approximately normally distributed if standardised by

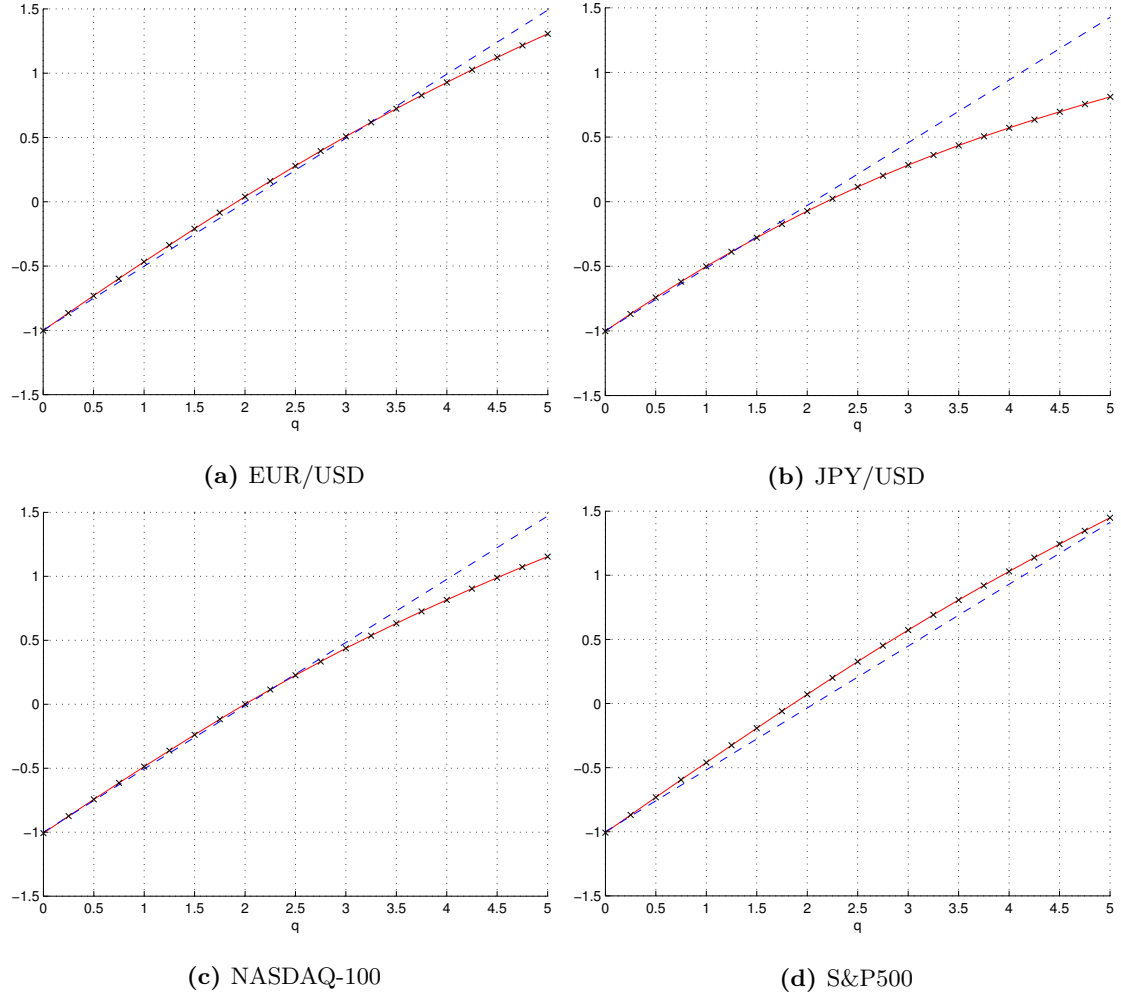


Figure 3.1: Estimated unifractal and multifractal scaling functions. Solid lines correspond to estimated scaling functions for the multifractal case (obtained using the partition function estimator) and dashed lines correspond to the estimates under the assumption of unifractality (obtained using the DMA estimator of Section 2.3.1).

their corresponding (time-varying) realised volatilities for each day. As in Chapter 2, a 5th order AR-RV(5) model was used since it was found to produce the best average density forecasting performance for the dataset employed here.

In addition to these two previous benchmark methods, one variant of the autoregressive unifractal (AR-UF) density forecasting model proposed in Chapter 2 was also added to the empirical exercise of the current chapter. The AR-UF method is implemented as described in Chapter 2, with the self-affinity index H estimated from the intraday data using the detrended moving average method of Alessio et al. (2002) and daily return densities estimated from the rescaled intraday returns using the non-parametric kernel density function estimator (denoted by AR-UFNP in Chapter 2). The simpler variant of the unifractal method is employed for which the autoregressive parameter values used to produce density forecasts are fixed rather than time varying, since this was typically found to maximise density forecasting performance.

For all benchmark density forecasting methods, the same rolling window estimation scheme as described in Section 3.3 was employed for producing density forecasts: the parameters of the models are estimated using an n -day rolling window of data (daily data in the case of the GARCH model and 5-minute intraday data for the case of the AR-RV model) and these parameter estimates are then used to produce one-step-ahead point forecasts for the relevant moments of daily returns. When combined with the relevant parametric form assumed for the return distribution, this allows one-step-ahead out-of-sample density forecasts to be produced for daily returns.

3.4.3 Methods for Density Forecast Comparison

The first method used for comparing the out-of-sample density forecasting performance of the methods is the same statistical test for equal predictive ability of Gneiting and Ranjan (2011) based on the continuous ranked probability score that was previously employed in Chapter 2. The test assumes that two competing forecasting methods

are used to produce one-step-ahead out-of-sample density forecasts for the variable of interest, y . It is assumed that N density forecasts are produced by each forecasting method and the forecasts produced by the two models at time t (for use at time $t + 1$) are denoted by $\tilde{f}_t(y)$ and $\tilde{g}_t(y)$, respectively.

The loss function employed by the test is the continuous ranked probability score (CRPS), generalised to allow more importance to be placed on forecast accuracy in particular regions of the density via the use of a weighting function. The value of the weighted CRPS for the forecast produced by each of the forecasting methods in period $t + 1$ is denoted by $S(\tilde{f}_t, y_{t+1})$ and $S(\tilde{g}_t, y_{t+1})$ respectively. The average value of the weighted CRPS in (4.1) can be calculated for each of the two density forecasting models over the N out-of-sample periods (for period $m + 1$ until period T) as:

$$\bar{S}^f = \frac{1}{N} \sum_{t=m}^{T-1} S(\tilde{f}_t, y_{t+1}) \quad \text{and} \quad \bar{S}^g = \frac{1}{N} \sum_{t=m}^{T-1} S(\tilde{g}_t, y_{t+1}) \quad (3.4.1)$$

A formal test can then be based on the following test statistic:

$$t = \frac{\bar{S}^f - \bar{S}^g}{\hat{\sigma}_n / \sqrt{N}} \quad (3.4.2)$$

where $\hat{\sigma}_n^2$ is a standard heteroskedasticity and autocorrelation consistent estimator for the asymptotic variance of $\sqrt{N}(\bar{S}^f - \bar{S}^g)$.

Under the null hypothesis that the two density forecasting models have equal predictive ability, the test statistic in (3.4.2) is asymptotically normally distributed, with the null rejected at the $\alpha\%$ significance level if $|t| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution. Given that lower values of the CRPS correspond to better forecasts, in the case of rejection, the forecasting model f should be chosen when the sample value of the test statistic is positive and model g when it is

negative.

The CRPS-based test above is a purely statistical measure of predictive ability and as such can give no indication of the economic gains or losses that would be realised by applying the various density forecasting methods in practice. Therefore, the CRPS-based test is supplemented by a second density forecast comparison method based on the problem of optimal portfolio allocation between a risky and a risk-free asset.

It is assumed that at time t an investor has total wealth of 1 to allocate between a single risky asset and a risk-free asset. The proportion invested in the risky asset is given by ω_t , with the remainder invested in the risk-free asset. Denoting the risky and risk-free returns from time t to $t+1$ by r_{t+1}^r and r_t^f respectively, the value of the portfolio at time $t+1$, denoted W_{t+1} , is then given by:

$$W_{t+1} = 1 + \omega_t r_{t+1}^r + (1 - \omega_t) r_t^f$$

The utility of the investor at time $t+1$ is assumed to depend on final wealth W_{t+1} according to a power utility function, with a coefficient of relative risk aversion γ :

$$\begin{aligned} U(W_{t+1}) &= \frac{W_{t+1}^{1-\gamma}}{1-\gamma} \\ &= \frac{1}{1-\gamma} \left[1 + \omega_t r_{t+1}^r + (1 - \omega_t) r_t^f \right]^{1-\gamma} \end{aligned}$$

When choosing ω_t at time t , r_{t+1}^r the rate of return on the risky asset from time t to $t+1$, is unknown and so the investor chooses the portfolio weight in order to maximise the expected utility obtained at $t+1$. Formally, the optimal weight ω_t^* at time t is obtained as the solution to:

$$\begin{aligned} \omega_t^* &= \arg \max_{\omega_t} E_t [U(W_{t+1})] \\ &= \arg \max_{\omega_t} E_t \left[\frac{1}{1-\gamma} \left[1 + \omega_t r_{t+1}^r + (1 - \omega_t) r_t^f \right]^{1-\gamma} \right] \end{aligned} \quad (3.4.3)$$

Rewriting (3.4.3) using the standard expression for the expectation of a random variable gives:

$$\omega_t^* = \arg \max_{\omega_t} \int \frac{1}{1-\gamma} \left[1 + \omega_t r_{t+1}^r + (1 - \omega_t) r_t^f \right]^{1-\gamma} \tilde{f}_t(r_{t+1}^r) dr_{t+1}^r \quad (3.4.4)$$

where, consistent with previous notation, $\tilde{f}_t(r_{t+1}^r)$ is the density forecast produced at time t for the risky return r_{t+1}^r . From (3.4.4) it is clear that different density forecasts for r_{t+1}^r will lead to different portfolio allocations at time t and therefore different realised utilities at time $t + 1$.

It is assumed that the investor holds the portfolio defined by the weight ω_t^* for a single period, before readjusting the portfolio weights based on new information in the following period. In the current empirical exercise this portfolio readjustment is performed daily and the portfolio allocation is made between one of the risky assets discussed in Section 3.4.1 and a risk-free asset, which is represented by the 3-month Treasury bill rate (with the rate converted to a daily return). Solving the portfolio allocation problem in equation (3.4.4) in each of the N days in the out-of-sample period results in a time series of portfolios, which in turn produces a time series of N realised utilities once the true risky return for the following period is observed.

The relative performance of the portfolios obtained from the density forecasting methods is then compared using the certainty equivalent return (CER) of the portfolio, which is defined as follows:

$$CER = \left[(1 - \gamma) \frac{1}{N} \sum_{t=1}^N RU_t \right]^{\frac{1}{(1-\gamma)}} - 1$$

where RU_t is the realised utility obtained from the portfolio in period t . The CER gives the risk free rate of return that would provide the same average level of realised utility as the portfolio over the out-of-sample period, implying that higher CER values are

preferable to lower values.

3.4.4 Empirical Results

For the empirical results presented in this section a 250-day rolling in-sample window (or value of n in the notation of Section 3.4.3) is used for parameter estimation and the density forecasts are compared over a 750 working day evaluation period, from the 250th until the 1000th working day in the sample for each series⁶.

Following Andersen et al. (2003), the initial order for the autoregressive component of the AR-MFVK(p, q) specification was set equal to 5 (i.e. one working week). The adequacy of these initial dynamic specifications was then checked by applying the standard Ljung-Box test for residual autocorrelation to the residuals obtained from fitting each time series model over the complete sample period. In almost all cases, for these initial 5th order models the null of no residual autocorrelation could not be rejected at any conventional significance level, suggesting that the 5th order specifications are adequate for modelling the dynamic structure of the estimated moment series. The only exceptions were for the two equity index series, where the null of no residual autocorrelation was rejected for the variance component (but not the kurtosis component) at the 10% level for the NASDAQ-100 and the 5% level for the S&P500.

A final issue of model specification that must be investigated for the multifractal method is the optimal choice of size for the window of intraday data used to produce each rolling estimate of the scaling function and prefactor (m in the notation of Section 3.3.2). Table 3.1 contains sample values for the simple unweighted CRPS-based test statistic comparing the predictive ability of the multifractal method using various window sizes against the different benchmark methods.

It can be seen from Table 3.1 that on average the optimal window size is around 15

⁶Because of the difference in trading days, the start and end dates of this period differ slightly for the two series: for the EUR/USD it spans 19th Dec 2007 - 11th Nov 2010 and 31st Dec 2007 - 21st Dec 2010 for the S&P500 data

Table 3.1: Sensitivity of predictive ability to changes in estimation window length. Values correspond to the sample values of the simple unweighted CRPS-based test statistic of Section 4.3. The CRPS-based test statistic is asymptotically normally distributed under the null of equal predictive ability and the test statistic is constructed such that significant negative values imply the multifractal method is superior to the benchmark model. See Section 3.4.3 for further details.

Window size	2 days	5 days	10 days	15 days	20 days	25 days	50 days
EUR/USD:							
GARCH(1,1) benchmark	4.28	0.75	-1.72	-2.23	-2.71	-2.75	-2.76
AR-RV(5) benchmark	5.45	3.13	0.22	-0.32	-0.93	-0.87	-0.56
AR-UF(5) benchmark	5.93	4.71	2.26	1.80	1.18	1.26	1.24
JPY/USD:							
GARCH(1,1) benchmark	3.03	1.58	0.39	-0.63	-0.66	-0.65	-0.59
AR-RV(5) benchmark	3.38	2.70	1.70	0.75	0.69	0.61	0.52
AR-UF(5) benchmark	3.39	2.22	0.83	-0.65	-0.70	-0.69	-0.57
NASDAQ-100:							
GARCH(1,1) benchmark	4.87	2.00	0.44	1.04	1.15	1.22	1.73
AR-RV(5) benchmark	5.37	2.62	0.32	1.07	1.14	1.16	1.60
AR-UF(5) benchmark	5.36	3.91	2.24	3.17	3.06	2.91	2.78
S&P500:							
GARCH(1,1) benchmark	4.59	1.79	0.53	0.74	0.82	1.11	1.97
AR-RV(5) benchmark	5.26	2.42	0.50	0.71	0.73	0.95	1.66
AR-UF(5) benchmark	5.51	3.98	2.41	2.68	2.74	2.83	3.11

working days across the four assets; 10 working days appears to be approximately optimal for the two equity index series and 20 days for the exchange rate series, with these values used for the remainder of the empirical exercise. Longer windows increase the number of intraday observations available, but do not result in an improvement in forecasting performance, presumably because older intraday data are no longer informative about the current properties of the return process. Equally, shorter windows reduce density forecasting performance, either because some degree of smoothing produces superior estimates of daily return moments⁷, or because of limited finite sample performance of the chosen partition function estimator.

⁷Noise in the observed intraday return process could make estimates calculated from short periods less informative about the true behaviour of the underlying process. This is an issue previously encountered in the realised volatility literature (see for example Andersen et al., 2003), where various methods have been proposed to mitigate the problem.

Table 3.2: Out-of-sample density forecast comparison using CRPS-based test statistic. The CRPS-based test statistic is asymptotically normally distributed under the null of equal predictive ability and the test statistic is constructed such that significant negative values imply the multifractal method is superior to the benchmark model. See Section 3.4.3 for further details.

Weighting Function	None	Centre	Left Tail	Right Tail
EUR/USD:				
GARCH(1,1) benchmark	-2.71***	-2.91***	-3.73***	-0.57
AR-RV(5) benchmark	-0.93	-1.57	-1.15	-0.35
AR-UF(5) benchmark	1.18	0.62	0.46	1.03
JPY/USD:				
GARCH(1,1) benchmark	-0.66	-0.12	-1.31	0.22
AR-RV(5) benchmark	0.69	0.65	-1.16	2.22**
AR-UF(5) benchmark	-0.70	-0.67	-1.75*	0.64
NASDAQ-100:				
GARCH(1,1) benchmark	0.44	0.34	-0.24	1.01
AR-RV(5) benchmark	0.32	0.98	0.15	0.41
AR-UF(5) benchmark	2.24**	2.12**	2.03**	1.81*
S&P500:				
GARCH(1,1) benchmark	0.53	0.29	-0.34	1.14
AR-RV(5) benchmark	0.50	1.14	-0.32	1.23
AR-UF(5) benchmark	2.41**	2.85***	1.69*	2.48**

Table 3.2 presents a comparison of density forecasting performance between the multifractal method and the benchmark models using the CRPS-based test outlined in Section 3.4.3. Considering first the more established GARCH and AR-RV benchmarks, it is clear that the density forecasts from the multifractal model perform well for the EUR/USD data, frequently providing highly statistically significant improvements in predictive ability over the GARCH benchmark method. Compared to the more competitive AR-RV benchmark utilising intraday data, the sample values for the EUR/USD data are generally negative, implying that the multifractal method provides superior predictive ability, but the gains are not large enough to be statistically significant. In addition, the gains in forecasting performance from the multifractal method appear to vary across the regions of the density function; the performance of the unifractal method is particularly strong in the centre and left tail of the EUR/USD return density, suggest-

ing that it should perform well in risk management applications, such as the calculation of Value at Risk or expected shortfall.

From the later sections of Table 3.2, the multifractal method is clearly less competitive with the GARCH and AR-RV benchmarks for the other return series than for the EUR/USD data, with the method unable to provide a statistically significant improvement in predictive ability over the benchmark methods. Nonetheless, in all but one case, the null of equal predictive ability cannot be rejected, implying that the multifractal method is again competitive with these benchmark established forecasting methods.

The final unifractal AR-UF benchmark model of Chapter 2 typically provides the strongest density forecasting performance of the three benchmark methods. For the equity index data in particular, the unifractal AR-UF method consistently provides gains in predictive ability over the new multifractal method that are highly significant across the whole domain of the density. However, for the exchange rate data the relative performance of the unifractal and multifractal methods is closer, with the null of equal predictive ability not rejected in most cases. These empirical findings suggest that when modelling and forecasting the distribution of equity returns, the ability to employ non-parametric specifications for the daily return density provided by the unifractal approach of Chapter 2 is more beneficial than the additional flexibility in distributional scaling properties permitted by moving from a unifractal to a multifractal context.

Differences in relative density forecasting performance across the different return series were also observed for the previous unifractal method of Chapter 2, where differences in the type of distributional scaling across the assets were identified as a potential cause, with the unifractal method appearing to perform best for series closest to the required unifractal scaling behaviour. However, the observed differences in relative density forecasting performance for the multifractal method cannot simply be due to the same differences in the type of distributional scaling, since the multifractal approach of the current paper should be valid for data exhibiting either unifractal or multifractal scaling

behaviour.

A possible alternative explanation is provided by differences in the strength, rather than the type, of distributional scaling across the different return series. This can be measured to some extent by considering the standard errors of the estimated values of $\tau(q)$ in the regressions of $\log S_q(T, \Delta t)$ on $\log(\Delta t)$ that follow from equation (3.3.1). Whilst these standard errors could be calculated for the scaling function estimates from the complete sample of data, given the rolling estimation method used to obtain the dynamic estimates of the scaling properties it is more relevant to calculate standard errors for estimates of $\tau(q)$ obtained from rolling windows.

Adjusting the lengths of the rolling windows employed to compensate for the difference in trading hours per day between the two types of asset⁸, there is indeed evidence of differences in the strength of distributional scaling across the assets. The rolling estimates of $\tau(q)$ for the EUR/USD data have the lowest standard errors on average over the sample period, followed by the JPY/USD data. The standard errors for both equity index series are approximately double those for the EUR/USD data, suggesting that they do indeed exhibit weaker distributional scaling than the exchange rate series.

We next investigate the extension to the standard multifractal method proposed in Section 3.3.4, which modifies the basic method by randomly shuffling the intraday data in each rolling window before estimating the daily return variance and kurtosis. As previously discussed, one potential problem is that each time the data are randomly shuffled different estimates of daily return moments will be obtained, making the density forecasts produced by the model stochastic. The solution proposed for this problem in Section 3.3.4 is to repeat the shuffling process numerous times, producing multiple multifractal estimates of the daily return moments for each window of intraday data, before taking an average of these estimates over all of the repetitions. However, this

⁸A rolling window of 20 days was used for the exchange rate series and 65 days for the equity index series. This is necessary to ensure that each rolling estimate is calculated from approximately the same number of intraday observations, minimising the effects of sample size on the standard errors.

solution will only be effective if the moment estimates averaged over the repetitions converge to a particular value as the number of repetitions is increased and so whether this holds in practice should be investigated.

Figure 3.2 contains plots of the average moment estimates obtained using 1 to 2500 repetitions of the shuffling process described above for the EUR/USD data⁹. To ensure that the exercise is as relevant for the current context as possible, the samples of intraday data used have the same length of 20 working days used previously to estimate the daily return moments for each trading day; 3 different 20-day windows were tested from arbitrary points in the sample, beginning on the 1st, 500th and 1000th trading days respectively.

From Figure 3.2 it can be seen that although the average moment estimates from the shuffled intraday data do not converge entirely to specific values as the number of repetitions increases (at least up to the maximum of 2500 repetitions considered here), they do typically converge to a narrow range of values. Whilst these figures represent just one possible realisation for each of the 3 arbitrarily chosen windows, the same pattern of convergence was observed for other windows of intraday data chosen from the complete sample of EUR/USD data and also more generally for data from the other 3 asset return series. Furthermore, in none of the cases from Figure 3.2 do the average moment estimates from the shuffled data converge to the estimated values obtained from the original ordered data, with the differences in many cases being substantial. This implies that at least some of the multifractal scaling present in the original ordered data is due to the second source of multifractality mentioned in Section 3.4 (small and large fluctuations of the process having different long-term correlations) and so this modification to the method should produce noticeable changes in the resulting density forecasts.

Having established that the modified version of the multifractal method should be

⁹Equivalent figures for the other series have been omitted to conserve space, but similar results are observed for each.

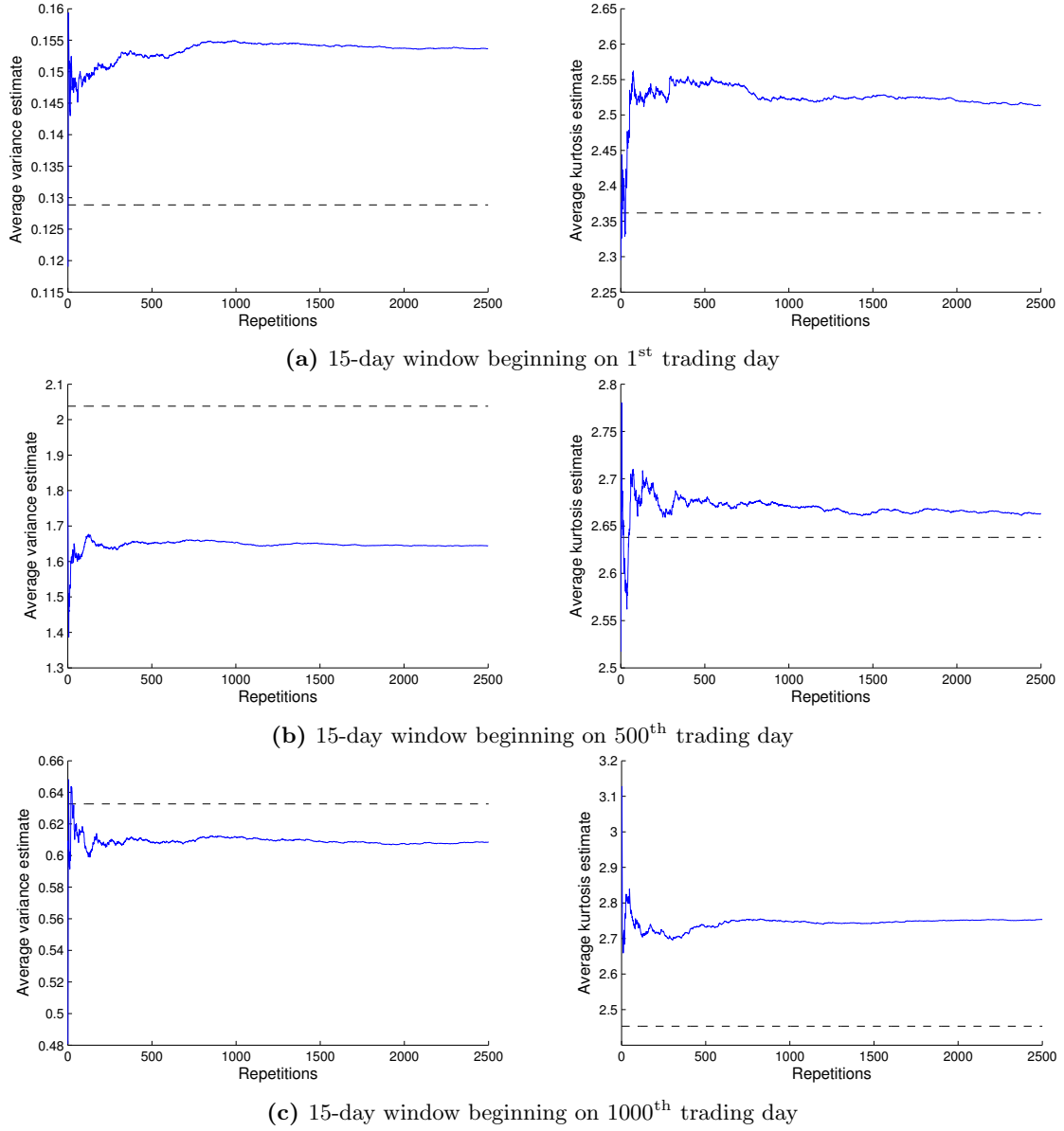


Figure 3.2: Multifractal moment estimates from shuffled and ordered intraday data
Dotted lines are multifractal moment estimates obtained from the original ordered 5-minute intraday data in the specified window. Solid lines are multifractal moment estimates obtained from the same intraday data shuffled randomly, averaged over the number of repetitions shown on the horizontal axis.

valid for the current dataset, the density forecasting performance of the method can now be investigated. Whilst Figure 3.2 shows that some variation may remain in the average moment estimates beyond 2500 repetitions, in practice it was found that repe-

tition numbers as low as 500 resulted in consistent density forecasting performance over the out-of-sample period¹⁰ and so as a compromise the number of repetitions was set to 1000.

Table 3.3 contains equivalent results for the modified multifractal method to those in

Table 3.3: Out-of-sample density forecast comparison using CRPS-based test statistic - shuffled data with 1000 repetitions used for multifractal approach

Weighting function	None	Centre	Left tail	Right tail
EUR/USD:				
GARCH(1,1) benchmark	-3.75 ***	-3.42***	-4.00 ***	-1.83*
AR-RV(5) benchmark	-2.78 ***	-2.47 **	-2.30**	-1.98 *
AR-UF(5) benchmark	-1.49	-1.02	-1.07	-0.93
JPY/USD:				
GARCH(1,1) benchmark	-0.27	0.21	-0.90	0.49
AR-RV(5) benchmark	1.16	1.29	-1.02	2.75***
AR-UF(5) benchmark	-0.18	0.05	-1.15	0.98
NASDAQ-100:				
GARCH(1,1) benchmark	-0.36	-0.07	-0.84	0.26
AR-RV(5) benchmark	-0.85	0.39	-0.68	-0.76
AR-UF(5) benchmark	1.23	1.43	1.27	0.61
S&P500:				
GARCH(1,1) benchmark	-0.45	-0.29	-0.64	-0.16
AR-RV(5) benchmark	-0.82	0.27	-0.69	-0.76
AR-UF(5) benchmark	1.19	1.73*	1.43	0.03

Table 3.2; from the sample values it can be seen that this modification to the multifractal method consistently improves the predictive ability of the multifractal method for the EUR/USD, NASDAQ-100 and S&P500 data. The changes in the sample values of the test statistics are often substantial in size, with the outcome of the test often changing as a result. Most notably, the unifractal AR-UF benchmark previously provided statistically significant improvements in predictive ability over the standard multifractal approach, but when compared to the modified multifractal approach the null of equal

¹⁰The remaining variation in the moment estimates for a given trading day will become less significant when comparing density forecasting performance in practice, since the CRPS-based test of equal predictive ability compares average forecasting accuracy over the complete length of the out-of-sample period (750-days in the current context).

predictive ability cannot typically be rejected at any conventional significance level. However, the same improvement in density forecasting performance is not found for the JPY/USD data for which the modification to the multifractal approach typically reduces performance, although only in the right tail of the return density are these changes large enough to alter the outcome of the test for equal predictive ability at any conventional significance level.

The improvements in predictive ability that are typically obtained from shuffling the data in this way suggest that of the two sources of multifractality highlighted in Section 3.3.4, multifractality due to a broad probability distribution is more relevant for the current application than that due to small and large fluctuations of the process having different long-range correlations. A possible explanation for this follows from the discussion and explanation of the multifractal method in Sections 3.3.1 to 3.3.3. It can be seen that at the point when the multifractal moment scaling law is applied to estimate the daily return moments, no dynamic structure has actually been imposed on the return process at any timescale; the dynamic structure required to produce forecasts for the moments (and ultimately the density function) of daily returns is imposed at a later stage onto the time series of estimated daily return moments. Intuitively therefore, it seems possible that the component of scaling that is due to the temporal structure of the ordered data is less relevant in the current application than the scaling of the unconditional distribution of returns at different timescales.

Finally, Table 3.4 contains the results for the portfolio allocation exercise discussed in Section 3.4.3. The reported values are the certainty equivalent returns (CER) expressed as an annualised percentage return for the expected utility maximising portfolio using the density forecasts from the GARCH and AR-RV benchmarks, plus the new multifractal AR-MFVK method¹¹. The portfolio allocation exercise has been performed with several

¹¹The unifractal approach of Chapter 2 has currently been omitted from this comparison, since the use of a non-parametric specification for the daily return density makes evaluating the integral of the density forecast in equation (3.4.5) more complex.

different values of the coefficient of relative risk aversion γ , in order to assess whether the optimal forecasting method varies with the level of investor risk aversion.

Table 3.4: Certainty equivalent returns from portfolio allocation exercise. Reported values are certainty equivalent returns (CERs) of the expected utility maximising portfolio for each forecasting method, with various levels of investor risk aversion. All CER values are expressed as annualised % rates of return.

Value of CRRA	1	2.5	5	10
EUR/USD:				
GARCH(1,1)	-9.29	-9.94	-10.31	-9.37
AR-RV(5)	-6.15	-6.29	-5.71	-4.84
AR-MFVK(5)	0.78	0.72	0.63	0.58
JPY/USD:				
GARCH(1,1)	-4.44	-3.65	-2.63	-3.02
AR-RV(5)	-1.38	-0.94	-0.15	0.23
AR-MFVK(5)	-0.27	-0.05	0.09	0.26
NASDAQ-100:				
GARCH(1,1)	27.07	23.95	17.34	8.99
AR-RV(5)	3.72	4.48	0.34	-1.15
AR-MFVK(5)	2.21	1.51	1.21	0.89
S&P500:				
GARCH(1,1)	23.04	18.24	10.72	6.01
AR-RV(5)	7.63	5.44	-0.92	-3.43
AR-MFVK(5)	1.57	1.20	0.99	0.78

From Table 3.4 it can be seen that the patterns observed when assessing density forecasting performance in the context of portfolio allocation are consistent with those previously observed in Table 3.2 in terms of the CRPS-based test statistic, thus reinforcing the previous empirical findings. For the exchange rate series the portfolios obtained from the multifractal approach provide the highest CER values, with those from the AR-RV benchmark in second place. For the EUR/USD series the gains from the multifractal approach over the benchmark methods are substantial, again confirming the previous finding of strong performance for the EUR/USD data. For the JPY/USD series the performance of the multifractal and AR-RV methods are typically closer, with

the largest differences observed at lower levels of risk aversion. For the equity index series the ranking of the forecasting methods is typically reversed, with the GARCH method producing portfolios with much higher CER values than either the AR-RV benchmark or the new multifractal method. It should however be noted the multifractal density forecasts produce portfolios with lower return volatility than the benchmark models and so for higher levels of risk aversion the multifractal method is actually able to provide higher a CER than the realised volatility approach.

The negative CER values observed in some cases may be due to the degree of risk aversion implicit in the investors' utility function, or the uncertainty around the density forecasts yielding portfolios with negative average rates of return. For the equity index series the first of these two factors is sufficient to explain all negative CER values, but for the exchange rate series both factors are relevant.

3.5 Conclusion

The current chapter has proposed a new method for estimating and forecasting the moments and probability density function of daily financial returns using intraday data. The method is based on a new application of results from the theory of multifractal processes that provide a formal statistical link between the moments of the return process at different sampling intervals, allowing the variance and kurtosis of daily returns to be estimated directly from high-frequency intraday data. In the current application, these moment estimates are incorporated into density forecasts of daily returns, however in other financial applications the variance and kurtosis of returns are also variables of substantial interest in their own right.

In principle, the incorporation of relevant information contained in the intraday data can provide gains when estimating daily return moments, compared to methods based purely on daily data. At the same time, in comparison to existing methods utilising

intraday data in the realised volatility literature, the multifractal approach preserves a greater proportion of the information contained in intraday returns by allowing the data to be used to directly estimate both the variance and kurtosis of daily returns. Compared to the autoregressive unifractal (AR-UF) density forecasting approach proposed in the previous chapter, this multifractal approach allows for more flexible distributional scaling of the return process across different sampling intervals. However, unlike the multifractal approach of the current chapter, the AR-UF approach allows the daily return density for each trading day to be modelled non-parametrically, thus enabling the intraday returns to directly influence all aspects of the daily return density and avoiding the need to choose a parametric distributional form for daily returns.

The predictive ability of density forecasts produced by the new multifractal method was compared to existing methods in an empirical application using 5-minute intraday data on Euro (EUR) and Japanese Yen (JPY) exchange rates against the US Dollar (USD) and the S&P500 and NASDAQ-100 equity indexes. For the EUR/USD data the multifractal method provides large improvements in predictive ability over the GARCH benchmark model and is competitive with existing realised volatility based methods. This strong performance is improved further when considering the modified multifractal method proposed in Section 3.3.5 using randomly shuffled observations from each window of intraday data; this modification further increases the existing gains in predictive ability over the GARCH benchmark and also allows the method to provide statistically significant improvements over the realised volatility based benchmark.

For the remaining asset return series, the density forecasting performance of the multifractal approach is competitive with the existing methods from the literature, with the null of equal predictive ability unable to be rejected in the majority of cases. As with the EUR/USD data, the modified multifractal approach using shuffled intraday data provides consistent improvements in predictive ability for both the S&P500 and NASDAQ-100 data; in this case, in no situations can the null of equal predictive ability

be rejected for the equity data, in contrast to the standard multifractal approach using ordered data for which the benchmark methods were found to be superior in some situations.

Finally, compared to the unifractal density forecasting method of Chapter 2, the predictive ability of the multifractal method is typically weaker; this is especially true for the equity index data, where the unifractal AR-UF method consistently provides highly significant gains in predictive ability over the multifractal method across the whole domain of the return density. This empirical finding suggests that in the case of equity index data, the ability to employ nonparametric specifications for the daily return density is more beneficial than the additional flexibility in distributional scaling properties permitted by moving from a unifractal to a multifractal context. For the exchange rate data however the relative performance of the unifractal and multifractal methods is substantially closer, with the null of equal predictive ability not able to be rejected in most cases.

These empirical findings are reinforced by the results of a portfolio allocation exercise, in which the density forecasts from the competing methods are employed to optimally allocate funds between a risky and risk-free asset. In this context it was again found that the new multifractal approach can provide substantial gains over existing methods for the EUR/USD data, when measured in terms of the certainty equivalent return of the resulting portfolio. For the other assets the multifractal approach is found to outperform the realised volatility based method for higher levels of investor risk aversion, although for the equity index series both of the intraday methods are outperformed by the GARCH benchmark method.

A possible explanation identified for this variation in forecasting performance across the various return series is provided by differences in the strength of distributional scaling for the return series; the EUR/USD data seem to exhibit much stronger distributional scaling than the other series, with the JPY/USD data having the weakest scaling and the

NASDAQ-100 and S&P500 data in between these extremes. Thus it seems that there is some positive relation between the strength of the distributional scaling exhibited by a given time series of data and the resulting density forecasting performance of the multifractal method.

There are several possible changes that could be made to the current implementation of the multifractal method that could potentially improve density forecasting performance further. The first is to identify an alternative parametric form for the daily return density that does not require restrictions to be placed on the daily return kurtosis, as is necessary with the current location-scale t -distribution. Secondly, more flexible dynamic specifications could be tested for modelling and forecasting the daily return moments to replace the simple autoregressive models currently used; given the ability of the multifractal approach to estimate moments of returns at any chosen timescale from the same intraday data, specifications employing data at different sampling intervals could be employed, such as the mixed data sampling (MIDAS) and heterogeneous autoregressive (HAR) models previously applied by Clements, Galvão, and Kim (2008) to the problem of producing quantile forecasts for daily returns from realised volatility measures.

Chapter 4

A Statistical Test for Unifractality Versus Multifractality

4.1 Introduction

As discussed in the preceding chapters, the key characteristic of both unifractal and multifractal processes is the presence of distributional scaling, such that the statistical properties of the process at different sampling intervals are formally linked according to some theoretical scaling laws. Both unifractal and multifractal processes have been previously employed in financial applications, with numerous theoretical models and estimation methods proposed for each class of process. However, the theoretical properties of the two classes of process differ, with the more general multifractal processes allowing for more flexible distributional scaling across timescales. This additional theoretical flexibility does however come at a cost, with problems such as parameter estimation, forecasting and simulation all more complex and computationally demanding than in the simpler and more restrictive unifractal case.

Given the tradeoff that exists between theoretical flexibility and practical complexity for these two classes of process, a natural question that arises is whether a specific sample of data is most consistent with a unifractal or multifractal process. This is an important issue when applying these processes to financial problems in general, but is especially relevant for the methods proposed in the earlier chapters of the current work, given that the type of process dictates which of the two forecasting methods is appropriate for a given dataset.

This question has been previously studied in the multifractal finance literature by the numerous empirical studies that aim to investigate the presence of distributional scaling in financial time series (see again the list of references in Chapter 1 for numerous examples). The majority of these studies have concluded that the return series for a wide range of financial assets appear to be more consistent with the class of multifractal processes than unifractal processes, though the strength of the evidence varies across different studies and assets.

The empirical analysis of these previous studies is however limited by the lack of a formal statistical test for distinguishing between unifractal and multifractal distributional scaling for a specific sample of data. This has necessitated the use of informal graphical testing procedures for distinguishing between the two types of process, with the standard approach being based on the scaling function described in earlier chapters. The key problem with this graphical approach is that it may imply multifractal scaling even when applied to simulated data from a purely unifractal process. This problem of ‘spurious multifractality’ has been noted and studied in the literature, with notable examples including the work of Lux (2004), Ludescher et al. (2011), Schumann and Kantelhardt (2011) and Grech and Pamuła (2012).

The issue of spurious multifractality is particularly problematic in smaller samples, making the existing informal graphical testing approach particularly unsuitable for testing the local scaling properties of the process over shorter sub-periods of the total sample. The topic of local distributional scaling properties is more relevant than global scaling for the work of the earlier chapters, in which scaling properties were studied in a dynamic context, allowing the parameters characterising the scaling structure to change over time. This may to some extent explain the nearly complete focus of previous empirical studies on testing the characteristics of the global rather than local scaling properties of the return process.

Despite the obvious limitations of the standard graphical method for distinguishing between multifractal and unifractal scaling, little effort has been made to develop improved tests that address these shortcomings and are robust to the problem of spurious multifractality. The most notable exception in the context of finance is the work of Lux (2004), in which a formal statistical test is proposed for testing the null hypothesis of ‘no multifractality’ versus the alternative of multifractality.

Although this method represents perhaps the first attempt to develop a formal statistical test for multifractal scaling, the methodology used implicitly imposes some re-

restrictions on the null and alternative hypotheses, which in turn limit the generality of the approach. Most importantly, the testing approach requires a specific theoretical multifractal process to be chosen under the alternative hypothesis, which means that it is not possible to consider general multifractality of an unknown or unspecified form. The method cannot therefore be viewed as a test for multifractality in general terms, but is instead a test for distributional scaling consistent with a specific multifractal process. Furthermore, the chosen theoretical multifractal process must nest a unifractal processes as a special case for certain parameter values (in addition to satisfying some additional constraints) and of course different choices of multifractal process from the set of suitable options may lead to different test outcomes.

Secondly, because of the simplistic resampling method employed to obtain the distribution of the relevant parameters under the null hypothesis of non-multifractal scaling, the test does not allow for serial dependence of any form (linear or non-linear) under the null. Given that the majority of unifractal processes may exhibit some form of serial dependence, the null hypothesis cannot be considered to correspond to unifractality in any general sense.

The only other formal testing procedure that appears to have been proposed in the literature is the approach of Wendt and Abry (2007) based on wavelet leaders, which was published in the signal-processing literature and has not been applied in the context of finance. The proposed methodology relies on a combination of wavelet-based methods to estimate the parameters of interest, combined with block bootstrap resampling to obtain the distributions of the relevant test statistics.

Crucially, unlike the earlier work of Lux (2004), this wavelet leader approach does indeed provide an explicit test for multifractality versus unifractality; however, in the proposed form the test imposes the constraint that the scaling function has a quadratic functional form under the alternative of multifractality, thus ruling out more general forms of multifractality. In addition, from the included Monte Carlo exercises examining

the properties of the test with simulated unifractal and multifractal processes, it is clear that the proposed wavelet-based method requires large sample sizes to perform well. Whilst this may not be problematic in some applications, such as those in the physical sciences, in financial applications the use of smaller samples may be required, due either to limited data availability or the desire to study the local scaling properties of a process over shorter sample periods.

In response to this gap in the literature, the current chapter develops a formal statistical testing framework for determining whether a given sample of data is most consistent with a unifractal or multifractal data generating process, which does not suffer from the same limitations as previous approaches. In particular, the proposed testing methodology is applicable generally, without making any assumptions concerning the form of multifractality under the alternative hypothesis and exhibits good performance in smaller samples under a wide range of conditions, thus permitting tests of both local and global scaling properties.

A set of possible test statistics are proposed for testing the null hypothesis of unifractality against the alternative of multifractality for a given sample of data and as with previous approaches, these statistics are based on differences that exist in the functions characterising the distributional scaling properties for unifractal and multifractal processes. Importantly however, the degree or strength of multifractality is measured non-parametrically via numerical differentiation of the estimates for these functions, thus avoiding the need to make any assumptions concerning the form of multifractality under the alternative hypothesis.

Due to the specific characteristics of the testing environment and the complex theoretical properties of unifractal and multifractal processes, the distributions of the proposed test statistics are non-standard and the relevant rates of convergence are unknown. It is shown that these difficulties can be overcome through the use of an appropriate model-based bootstrap resampling scheme, allowing the distributions of the test statis-

tics under the null of unifractality to be approximated in order to calculate critical values or p-values for the tests.

The proposed testing methodology is applied to simulated unifractal and multifractal data in an extensive series of Monte Carlo exercises, where it is shown to have good empirical size and power properties in wide range of situations. In particular, the power of the tests is shown to be robust against various forms of multifractality under the alternative and the tests perform well for sample sizes that would be considered as small in the multifractality literature, thus confirming the suitability of the methodology for the study of both local and global scaling properties. This is demonstrated in an empirical exercise in which the testing methodology is applied to study the local scaling properties of the intraday dataset used in previous chapters containing a selection of key financial assets.

The structure of the current chapter is as follows. Section 4.2 briefly reviews the theoretical properties of unifractal and multifractal processes that are relevant for the work of the current chapter, describes the methods used to estimate the relevant scaling properties and discusses the standard informal graphical testing approach previously employed in financial applications. Section 4.3 summarises the testing problem and presents the set of statistics for testing the null of unifractality versus multifractality, with Section 4.4 then discussing the model-based bootstrap resampling scheme proposed for obtaining the distributions of these statistics under the null. Sections 4.5 and 4.6 contain the Monte Carlo and empirical exercises respectively and finally, Section 4.7 concludes.

4.2 A Review of Unifractal and Multifractal Processes

The proposed methodology for testing whether a given sample of data is most consistent with a unifractal or multifractal data generating process is based on the distributional

scaling properties that are specific to these classes of process. Therefore, in order to provide a clear basis for the proposed testing procedure, Section 4.2.1 begins with a brief review of the relevant theoretical properties of unifractal and multifractal processes¹.

In practical situations where the true data generating process for a given sample of data is unknown, such as the testing problem considered here, the parameters characterising the distributional scaling properties of interest must be estimated from the available data. Numerous methods have been proposed in the literature for this task and Section 4.2.2 provides a detailed discussion of the specific estimators that will be employed later during the paper. Finally, Section 4.2.3 discusses the informal graphical testing methodology previously used in the finance literature and the potential problems associated with this approach for distinguishing between unifractal and multifractal scaling.

4.2.1 Theoretical Scaling Properties

The fundamental property of both unifractal and multifractal processes is the presence of *distributional scaling* or *scale invariance*. On an intuitive level, this implies that the behaviour of the process observed at one timescale or sampling interval is, after an appropriate transformation, identical in a statistical sense to that observed at another timescale. Whilst both unifractal and multifractal processes possess this distributional scaling property, the structure of scaling differs between the two classes of process, with the latter permitting a more flexible relationship between the properties of the process at different timescales.

For both multifractal and unifractal processes, it can be shown more formally (see for example Mandelbrot et al., 1997) that a stochastic process X_t with increments $X_{t+\Delta t} - X_t$

¹More detailed treatments of these topics can be found in Mandelbrot et al. (1997), Calvet and Fisher (2002) or Kantelhardt (2009).

is unifractal or multifractal² if these increments are stationary and satisfy:

$$E[|X_{t+\Delta t} - X_t|^q] \propto \Delta t^{\tau(q)+1} \quad \text{or equivalently} \quad E[|X_{t+\Delta t} - X_t|^q] \propto \Delta t^{qH(q)} \quad (4.2.1)$$

for all sampling intervals or timescales, Δt , and all q . The function $\tau(q)$ is the same scaling function from previous chapters and the function $H(q)$ is known as the *generalised Hurst exponent*. Both of the representations in equation (4.2.1) are employed in the literature, but are completely equivalent, with the generalised Hurst exponent and scaling function related via the identity $\tau(q) \equiv qH(q) - 1$.

The functions $H(q)$ or $\tau(q)$ characterise the distributional scaling structure of the underlying process X_t , by formally describing how the moments of the absolute increments of X_t scale with the sampling interval Δt . In financial applications the process X_t is typically taken to be the logarithmic price process for a particular financial asset, which implies that equation (4.2.1) describes how the moments of the absolute returns scale with changes in the return timescale Δt . Numerous empirical studies, such as those previously cited in Chapter 1, have confirmed the presence of distributional scaling consistent with equation (4.2.1) in the return series for a wide range of financial assets. More recently however, the application of multifractal models in finance has been extended to model other characteristics of financial time series, such as the application to inter-trade duration of Chen et al. (2013).

The requirement that the increments $X_{t+\Delta t} - X_t$ in equation (4.2.1) are stationary does not have to hold for the original time series of interest and unifractal or multifractal time series may be non-stationary. However, in this case an appropriate transformation

²Alternative representations, such as those presented in earlier chapters, may also be employed to define the scaling behaviour of unifractal or multifractal processes, but the above definition is the most relevant for the current chapter.

must first be applied to the series that results in stationary increments, before the distributional scaling relationships of equation (4.2.1) will hold. In practice, as will be seen in Sections 4.2.2 and 4.2.3, the majority of estimators for $H(q)$ and $\tau(q)$ commonly employed in the literature perform some form of local detrending in order to remove non-stationarities.

The distributional scaling law in equation (4.2.1) holds for both multifractal and unifractal processes and for both types of process $H(q)$ and $\tau(q)$ share some common properties. Firstly, from equation (4.2.1) at $q = 0$ it is clear that $\tau(0) = -1$ and thus the intercept of the scaling function is identical for both unifractal and multifractal processes. Secondly, $H(2)$ is directly connected to the dependence properties of the increment series; if the increments are independent $H(2) = 0.5$, if the increments are persistent then $H(2) > 0.5$ and finally if the increments are anti-persistent then $H(2) < 0.5$.

Despite these similarities, there are important differences in the functional forms for $H(q)$ and $\tau(q)$ between unifractal and multifractal processes. Specifically, it can be demonstrated (see for example Calvet and Fisher, 2002) that the following theorem holds:

Theorem 4.1: *For a multifractal process, the scaling function $\tau(q)$ is a strictly concave function of q and for a unifractal process $\tau(q)$ is a linear function of q . Equivalently, for a multifractal process the generalised Hurst exponent $H(q)$ is a non-constant function of q and for a unifractal process $H(q)$ is constant and independent of q .*

In the multifractal case, the conditions in Theorem 4.1 that must be satisfied by the generalised Hurst exponent or the scaling function clearly do not define unique functional forms for $\tau(q)$ or $H(q)$ for the class of multifractal processes as a whole. However, various theoretical multifractal processes have been developed in the literature for which $\tau(q)$ and $H(q)$ can be derived as functions of one or more of the parameters of the specific process. For example, for the multifractal random walk (MRW) of Bacry et al. (2001)

previously employed in the finance literature³ by Muzy et al. (2001), the scaling function and thus the generalised Hurst exponent are completely characterised by the parameter λ via:

$$\tau(q) = -(\lambda^2/2)q^2 + (1/2 + \lambda^2)q - 1$$

For the MRW, the limiting value of $\lambda^2 = 0$ results in a linear unifractal scaling function and thus λ^2 controls the strength of multifractality for the process, with estimated values in the range $0.02 \leq \lambda^2 \leq 0.035$ typically obtained for financial data in previous empirical studies (see for example Muzy et al., 2001). Figure 4.1 plots the theoretical $\tau(q)$ and $H(q)$ functions for a MRW process with several values of λ^2 , demonstrating the convergence of the MRW towards a unifractal process as the multifractality strength parameter $\lambda^2 \rightarrow 0$.

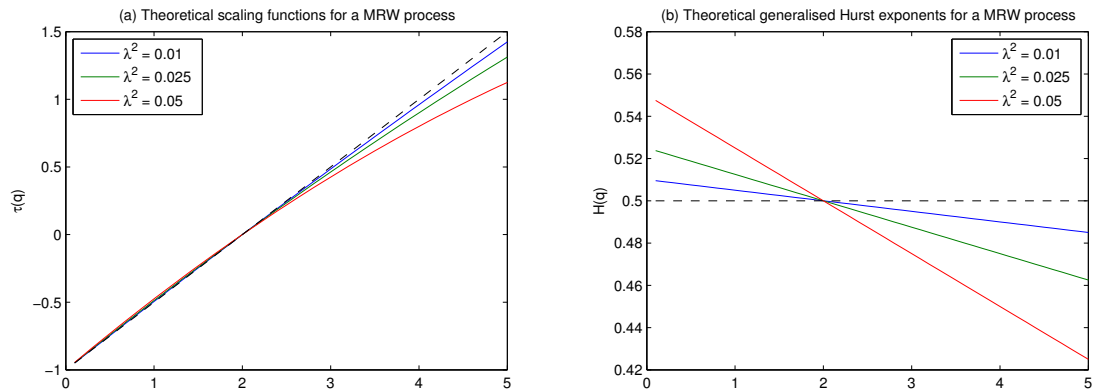


Figure 4.1: Theoretical values of the scaling function and generalised Hurst exponent for a log-normal multifractal random walk process with uncorrelated increments for various values of the parameter λ^2

The specific MRW process of Figure 4.1 is constructed to have a linearly independent increment series and thus $H(2) = 0.5$ for all values of λ^2 in Figure 4.1(b). The MRW

³Other examples of multifractal processes previously employed in financial applications include the multifractal model of asset returns of Mandelbrot et al. (1997) and Calvet and Fisher (2002) and the Markov-switching multifractal model of Calvet and Fisher (2004) and Calvet et al. (2006).

process can be generalised to allow for linear dependence in the increment series, however this is not typically required in the context of financial return series since the process allows for non-linear serial dependence even when linear dependence is assumed. This allows the MRW to reproduce the volatility clustering observed in financial returns without requiring the level of the returns to be correlated.

In the unifractal case, from Theorem 4.1 $H(q) = H \forall q$ and from the identity above relating $H(q)$ and $\tau(q)$ it follows that the linear unifractal scaling function is of the form $\tau(q) = Hq - 1$. Therefore, the single scalar parameter H completely defines the distributional scaling properties of a unifractal process, implying that all unifractal process with the same simple Hurst exponent H will share the same theoretical scaling properties with identical generalised Hurst exponents and scaling functions. This is in contrast to the class of multifractal processes, for which the functional forms for $H(q)$ and $\tau(q)$ differ from one multifractal process to another.

The constant value of H is referred to as *the* Hurst exponent or the *simple* Hurst exponent⁴, in order to distinguish it from the generalised Hurst exponent $H(q)$ in the multifractal case. Furthermore, as previously discussed the persistence properties of the increments of the process for both unifractal and multifractal processes are determined by the value of $H(2)$ and as a result, the value of the simple Hurst exponent also determines the persistence properties for the increment series, in addition to the distributional scaling properties.

Figure 4.2 plots the theoretical values of the generalised Hurst exponent and the scaling function against q for a generic unifractal processes for various values of the Hurst exponent, H ; as discussed above, $\tau(q)$ is linear with slope H and intercept -1 and $H(q)$ is constant and equal to H .

⁴The term *self-affinity index* was used in the earlier chapters and is also widely employed in the literature on unifractal processes, however the Hurst exponent terminology will be employed during the current chapter to remain consistent with the multifractal case.

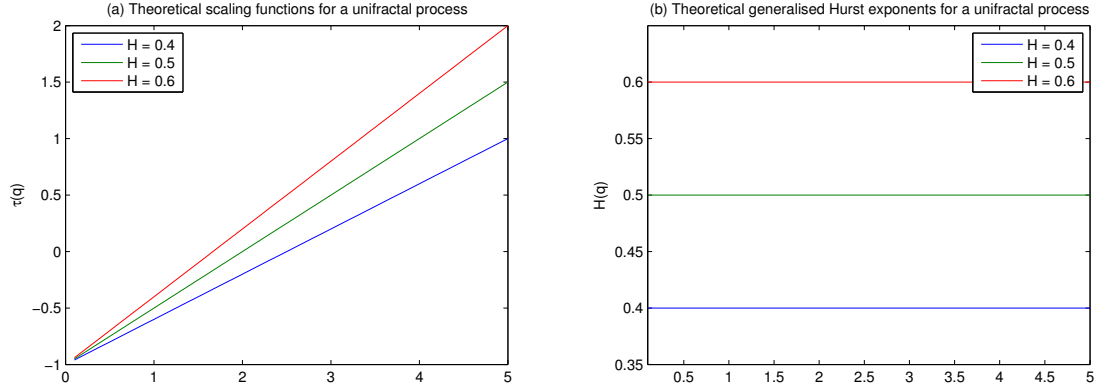


Figure 4.2: Theoretical values of the scaling function and generalised Hurst exponent for a generic unifractal processes for various values of the simple Hurst exponent, H

4.2.2 Estimation of Scaling Properties

The differences between the theoretical scaling properties of unifractal and multifractal processes discussed above provide a possible basis for testing whether a given sample of data is most consistent with a unifractal or multifractal data generating process. In practice however the true theoretical generalised Hurst exponent or scaling function of a given process is unknown and so the first step in the implementation of such a test is to estimate $H(q)$ and $\tau(q)$ from the available data.

In the unifractal case, $H(q)$ and thus $\tau(q)$ are completely characterised by the simple Hurst exponent H and so estimation of the scaling properties of a unifractal process requires the estimation of just a single scalar parameter. The most common estimators employed in this context include methods based on local detrending, such as the detrended fluctuation analysis and centred moving average estimators, and methods based on wavelet analysis, such as the discrete wavelet transform approach. A survey of these and other common estimators for the unifractal case can be found in the survey by Kantelhardt (2009), however estimation methods for unifractal processes will not be discussed separately here in detail. This is because the specific multifractal estimators employed in the current paper are generalisations of previous estimators developed

for the unifractal context that nest the corresponding unifractal estimators as special cases. As a result, the derivation and theory of the original unifractal estimators follows from the discussion of their multifractal generalisations and does not require separate exposition.

The specific methods employed for the multifractal case to estimate the functions $H(q)$ and $\tau(q)$ in the current paper are the multifractal detrended fluctuation analysis estimator of Kantelhardt et al. (2002) and the multifractal centered moving average estimator of Schumann and Kantelhardt (2011). The simpler partition function estimation approach is frequently employed in the finance literature (see for example Calvet and Fisher, 2002), but unlike the estimators above, is only strictly valid for stationary and normalised series. Whilst this condition is likely to be satisfied by financial returns, attention is restricted to estimators that are valid for stationary and non-stationary series to ensure that the testing approach developed here is as widely applicable as possible. Alternative estimators proposed in the literature for the multifractal context include the wavelet transform modulus maxima (WTMM) method of Muzy et al. (1991) and the later wavelet leader approach of Jaffard et al. (2007).

Multifractal Detrended Fluctuation Analysis Estimator

The multifractal detrended fluctuation analysis (MF-DFA) estimator of Kantelhardt et al. (2002) was developed as a generalisation of the earlier detrended fluctuation analysis (DFA) estimator for the unifractal context, providing an estimate of the generalised Hurst exponent $H(q)$ for any given value of q for a multifractal process as opposed to the simple (scalar) Hurst exponent in the unifractal case.

Suppose that the series of interest is denoted x_t and is of length T . In standard financial applications, x_t should be chosen as the return series of the asset of interest rather than the price series, since the series is cumulatively summed in the first stage of estimation. Alternatively, the price series may be used and the cumulative summation

in the first step below may be omitted. The MF-DFA estimate of the generalised Hurst exponent is then obtained as follows:

1. Calculate the cumulative sum or ‘profile’ series $\{y_t\}_{t=1}^T$ as $y_t \equiv \sum_{s=1}^t x_s$. At this stage the mean value of x_t is sometimes subtracted before cumulatively summing the observations, but this is not required since the series will automatically be demeaned in stage 3 below.
2. Divide the profile series y_t into N_s non-overlapping segments of equal length s . Since T will not generally be exactly divisible by s , the number of segments is given by the integer part of T/s , discarding some observations at the end of the series.⁵
3. The local polynomial trend of order m is calculated for y_t over each of the N_s segments, with the value of the fitted polynomial trend for the i -th observation in segment $\nu = 1, \dots, N_s$ denoted by $\tilde{y}_{\nu,i}$. For each of the N_s segments the local polynomial trend is then used to detrend the profile series and the variance of the detrended series over each segment is then calculated as:

$$F_{s,\nu}^2 = \frac{1}{s} \sum_{i=1}^s \{y_{[(\nu-1)s+i]} - \tilde{y}_{\nu,i}\}^2$$

for $\nu = 1, \dots, N_s$. The higher the order of polynomial detrending, m , the more complex the forms of nonstationarity that will be removed from the series. However, it is common in the literature to employ simple linear detrending ($m = 1$) due to the computational demands of higher order detrending.

4. Obtain the q -th order fluctuation function from the N_s variance terms:

⁵If desired, to avoid discarding observations steps (1) and (2) can be repeated starting from the end of the original series x_t , producing $2N_s$ segments of length s . However, the effects on the resulting estimates of $H(q)$ will typically be small unless s is large relative to T , which is in turn likely to produce unreliable estimates.

$$F_{q,s} = \left\{ \frac{1}{N_s} \sum_{\nu=1}^{N_s} [F_{s,\nu}^2]^{q/2} \right\}^{1/q}$$

where the order q is not constrained to take integer values, but can take any real value except for 0.

5. For a unifractal or multifractal process the q -th order fluctuation functions satisfy the following relationship:

$$F_{q,s} \propto s^{H(q)} \quad \text{or equivalently} \quad \log F_{q,s} = a + H(q) \log s$$

where $H(q)$ is the generalised Hurst exponent and, following the discussion in Section 4.2.1, for a unifractal process $H(q)$ is constant and independent of q , whereas for a multifractal process $H(q)$ varies with q . Steps 2 to 4 are repeated for different values of the segment size or scale, s , between some minimum and maximum values s_{min} and s_{max} . The value of $H(q)$ for the chosen value of q can then be estimated from a linear fit of $\log F_{q,s}$ against $\log s$.

6. Repeating steps 4 and 5 for various values of the order q produces a set of estimated values for the generalised Hurst exponent, $H(q)$. If required, an estimate of the corresponding scaling function, $\tau(q)$, can be obtained from the estimate of $H(q)$ via the identity $\tau(q) \equiv qH(q) - 1$ previously stated in Section 4.2.1.

Finally, it should be noted that for the value of $q = 2$ the multifractal DFA estimator simplifies to the earlier DFA estimator developed for estimating the scalar simple Hurst exponent, H , in the unifractal case.

Whilst there is no formal method for choosing the minimum and maximum segment sizes or scales, s_{min} and s_{max} , previous studies using simulated unifractal and multifractal processes (see for example Bashan et al., 2008) have found that the inclusion of

either very small scales or scales that are large relative to the sample size generally lead to higher average estimation error for $H(q)$. For large values of s , this occurs because very few non-overlapping segments can be obtained from the profile series y_t and so the data can provide little information concerning the properties of the process at that scale. For very small scales the low number of observations within each segment make accurate fitting of the local polynomial trends in step 3 difficult, again leading to unreliable estimates of the properties of the process at that scale. These previous studies have suggested approximate limits for the optimal minimum and maximum scales that are in the ranges of $5 < s_{min} < 15$ and $T/10 < s_{max}T/5$; these values were used as a starting point for the Monte Carlo studies of Section 5 and were typically found to perform well in the current context.

Multifractal Centred Moving Average Estimator

In a similar manner to the MF-DFA estimator of the previous subsection, the multifractal centred moving average (MF-CMA) estimator of Schumann and Kantelhardt (2011) was developed as a multifractal generalisation of an earlier estimator for the simpler unifractal context, namely the centred moving average (CMA) estimator of Alvarez-Ramirez et al. (2005).

As above for the MF-DFA estimator, denote the series of interest by x_t for $t = 1, \dots, T$, which in financial applications should be chosen as the return series of the asset of interest. The MF-CMA estimate of the generalised Hurst exponent $H(q)$ is then obtained as follows:

1. Calculate the cumulative sum or ‘profile’ series $\{y_t\}_{t=1}^T$ as $y_t \equiv \sum_{s=1}^t x_s$.
2. A centred moving average of the profile series, denoted $\bar{y}_{s,t}$ is obtained for the window length s via a moving average filter of the form:

$$\bar{y}_{s,t} = \frac{1}{s} \sum_{j=-(s-1)/2}^{(s-1)/2} y_{t+j}$$

for $t = 1 + (s-1)/2, \dots, T - (s-1)/2$. Note that the length of the moving average window s must be an odd number and the moving average filtered series is not defined for the first and last $(s-1)/2$ observations of the original series⁶.

3. As with the MF-DFA estimator above, the profile series y_t is split into N_s non-overlapping equal length segments of length s , where s is the same moving average window size used above in stage 2. Again, the loss of some observations at the end of the series can be avoided if desired by repeating the same process starting from the end of the series, thus producing $2N_s$ segments of length s . Within each of the segments the profile series is detrended using the moving average trend and the variance of the detrended observations over each segment is calculated as:

$$F_{s,\nu}^2 = \frac{1}{s} \sum_{i=1}^s \{y_{\nu s+i} - \bar{y}_{s,\nu s+i}\}^2$$

for $\nu = 0, \dots, N_s - 1$.

4. Obtain the q -th order fluctuation function from the N_s variance terms:

$$F_{q,s} = \left\{ \frac{1}{N_s} \sum_{\nu=0}^{N_s-1} [F_{s,\nu}^2]^{q/2} \right\}^{1/q}$$

where the order q is not constrained to take integer values, but can take any real value except for 0.

5. For a unifractal or multifractal process the q -th order fluctuation functions satisfy the following relationship:

⁶Schumann and Kantelhardt (2011) propose methods for including these edge observations in the estimation procedure, however the differences will be small unless the value of s is large relative to the sample size.

$$F_{q,s} \propto s^{H(q)} \quad \text{or equivalently} \quad \log F_{q,s} = a + H(q) \log s$$

where $H(q)$ is the generalised Hurst exponent. Steps 2 to 4 are repeated for different values of the segment size, s , between some minimum and maximum values s_{min} and s_{max} . The value of $H(q)$ for the chosen value of q can then be estimated from a linear fit of $\log F_{q,s}$ against $\log s$.

6. Repeating steps 4 and 5 for various values of the order q produces a set of estimated values for the generalised Hurst exponent, $H(q)$. If required, an estimate of the corresponding scaling function, $\tau(q)$, can be obtained from the estimate of $H(q)$ via the identity $\tau(q) \equiv qH(q) - 1$ previously stated in Section 4.2.1.

It can be seen above that the implementation of the MF-CMA estimator is very similar to that of the MF-DFA estimator, differing only during steps 1 and 2 where the detrending of the profile series is performed using a moving average, rather than the series of local polynomial fits used by MF-DFA. This change results in an estimator that provides similar performance to the MF-DFA approach with simple linear detrending, but is less computationally demanding. The MF-CMA estimator cannot however be extended in the same way as MF-DFA (by using higher order polynomial fits within each segment) in order to account for higher order trends or non-stationarities present in the data.

As with the MF-DFA estimator, the corresponding CMA estimator previously developed for the simple Hurst exponent, H , in the unifractal case is obtained as a special case of the MF-CMA estimator for $q = 2$. The use of window sizes or scales, s , that are either very small or very large relative to the sample size will typically result in larger average estimation error for $H(q)$ for the same reasons discussed previously in the context of the MF-DFA estimator. Indeed, the same rules of thumb for selecting s_{min} and s_{max} given for the MF-DFA estimator in the previous subsection are also generally

appropriate for the MF-CMA estimator.

4.2.3 Previous Graphical Tests for Scaling & Spurious Multifractality

From Theorem 4.1, strict concavity of $\tau(q)$ or dependence of $H(q)$ on q implies a multifractal process and conversely linearity of $\tau(q)$ or independence of $H(q)$ on q implies a unifractal process. Previous empirical studies in the finance literature have employed informal graphical methods based on this principle in order to check for the presence of multifractal distributional scaling in financial data (see for example Muzy et al., 2001, Calvet and Fisher, 2002 or Di Matteo, 2007). In the finance literature this graphical check is performed almost exclusively using the scaling function representation of scaling behaviour rather than the generalised Hurst exponent representation, however the two should be equivalent.

To provide an example of this graphical testing approach, the estimated scaling functions and generalised Hurst exponents obtained for a selection of financial assets are plotted below in Figure 4.3. The dataset employed for Figure 4.3 is the same as that used for the empirical exercise of Section 4.6 and consists of 5-minute intraday returns spanning the period January 2007 to December 2010 for the EUR/USD and JPY/USD exchange rates and the NASDAQ-100 and S&P500 equity indexes, which provides large sample sizes of approximately 300,000 observations for the exchange rate series and 80,000 observations for the equity index series. To maintain consistency with the estimation methods for $H(q)$ and $\tau(q)$ used elsewhere in the current chapter, the MF-DFA estimator of Section 4.2.2 with linear local detrending is employed, rather than the simpler partition function estimator used previously for Figure 2.1 of Chapter 2. Furthermore, a wider range of segment sizes are used for estimation in Figure 4.3, with a minimum segment size of 5 and maximum segment sizes of 7,200 and 2,250 used for the exchange rate and equity index series respectively, in order to consider the scaling properties of the return series from 5-minute to approximately monthly sampling intervals.

The dashed lines are included for reference and represent the estimates obtained under the assumption of unifractality using the equivalent DFA estimator.

It is clear from Figure 4.3 that all of the estimated scaling functions are strictly concave and all of the estimated generalised Hurst exponents exhibit some degree of dependence on q . In previous empirical studies applying this graphical procedure to financial data, any observed concavity of $\tau(q)$ or dependence of $H(q)$ on q has been interpreted as a confirmation of multifractal distributional scaling in the series of interest. Following this approach, the estimated scaling functions and generalised Hurst exponents in Figure 4.3 lead to the conclusion that all series exhibit multifractal rather than unifractal distributional scaling behaviour.

If the true scaling properties of the process were observable then this would be a valid conclusion to draw, since by definition from Theorem 4.1 even a slight degree of concavity in $\tau(q)$ or dependence of $H(q)$ on q implies that the process in question is technically multifractal rather than unifractal (although for cases of very weak multifractality there may be little difference from unifractal behaviour). However, in practice these true scaling properties are not observable and analysis must be based on the estimates of $\tau(q)$ and $H(q)$ such as those in Figure 4.3.

Although some previous studies of estimator properties in the multifractal case have been performed using simulated processes (see for example Lashermes et al., 2005 and Schumann and Kantelhardt, 2011), the formal statistical properties of these estimators, such as finite sample biases and rates of convergence, have not been derived⁷. Nonetheless, it is straightforward to demonstrate informally that standard estimators can produce strictly concave estimates of $\tau(q)$ and non-constant estimates of $H(q)$, as expected for a multifractal process, even when applied to simulated data generated by a purely unifractal process.

⁷The theoretical properties of multifractal processes are complex and the task is further complicated by the additional choices that must be made to implement each of the estimators, including the domain of q over which estimation is performed and the segment sizes for the MF-DFA and MF-CMA estimators.

This issue of ‘spurious multifractality’ was first noted some time ago in the context of multifractal scaling in finance (see in particular Bouchaud et al., 2000 and Lux, 2004) and subsequently studied in more detail by work such as Grech and Pamuła (2012). The fundamental cause of this problem is the potential estimation error encountered when applying the estimators for the generalised Hurst exponent to finite samples of data.

More specifically, for a given sample size more information is available concerning the behaviour of the process at small timescales than at long timescales and as a result the properties at longer timescales can be estimated less precisely. Depending on the properties of the specific sample of data, the estimates of $H(q)$ obtained may suggest a more complex scaling relationship between the sampling interval and the properties of the process⁸ than is actually present in the data. Therefore, even when the true data generating process is purely unifractal, the effects of estimation error can make the scaling properties of the process appear consistent with those of the more general class of multifractal processes. As such, this problem of spurious multifractality invalidates the previous informal graphical testing procedures used in the finance literature and reinforces the need for a formal statistic test for unifractal versus multifractal scaling.

This problem typically becomes less severe as the sample size increases and so in principle it can at least be minimised by using a larger sample of data. However, there will be situations encountered in financial empirical applications where obtaining a sufficiently large sample may not be possible. For example, if high-frequency data are unavailable then a large sample of data can only be obtained by including data observed over a very long time period, in which case the observations from the beginning of the sample period may not be informative about the current properties of the process. Alternatively even if high-frequency data are available for the asset of interest, it may be the local rather than global properties of the return process that are of interest (as in Chapters 2 and

⁸Although $H(q)$ and $\tau(q)$ are explicit functions of the fluctuation function order q (see stage 4. of the MF-DFA and MF-CMA estimation algorithms of Section 4.2.2), implicitly they describe how the properties of the process at one sampling interval or timescale are related to those at other timescales.

3), necessitating the use of shorter rolling estimation windows and again limiting the available sample size.

As an illustration of this problem, Figure 4.4 contains the estimated generalised Hurst exponents and scaling functions for 3 realisations from a purely unifractal fractional Gaussian noise process of length of $T = 10,000$ in Figure 4.4(a) and $T = 2,500$ in Figure 4.4(b). As with Figure 4.3, the estimates are obtained using the MF-DFA estimator of Section 4.2.2, with minimum and maximum segment sizes of 5 and $T/10$ respectively. From Figure 4.4(a) the deviation from unifractal scaling is difficult to detect graphically from the sample estimates of $\tau(q)$ with series of length 10,000, despite being visible from the estimates of $H(q)$. However, when the series length is shortened to 2,500 the degree of spurious multifractality becomes apparent from both the sample estimates of the scaling function and the generalised Hurst exponent, with the former showing visible concavity and the latter showing clear dependence on q , as would be expected for a multifractal process.

It should also be noted that some other potential sources of spurious multifractality have been identified in the literature; in contrast to the issue of estimation error discussed above, which may occur even when the true data generating process is purely unifractal, these other sources are associated with situations in which the underlying process is neither purely unifractal nor multifractal. Most notably, the work of Ludescher et al. (2011) and Schumann and Kantelhardt (2011) demonstrated that spurious multifractality can also be found in processes with certain types of periodic trend, short-range dependence or structural breaks that become unifractal after being transformed in an appropriate manner to remove these effects. Whilst the proposed testing procedure may be applicable in these situations, they are not explicitly considered from this point onwards and are left as a possible direction for future research. This is because the specific aim of the current work is to develop a test for distinguishing between unifractal and multifractal scaling properties, but in their original untransformed form such processes

are not technically either unifractal or multifractal.

The one exception that will be considered further is the finding of Schumann and Kantelhardt (2011) that spurious multifractality may also be found for a time series that is locally unifractal within distinct sub-periods, but with the unifractal scaling exponents varying from one sub-period to another. As briefly mentioned in Chapter 2, such a process is often referred to as ‘multifractional’ in the literature (distinct from multifractal) and given the dynamic estimation framework employed in Chapter 2, the assumption of unifractality within each estimation window imposed there really corresponds to the assumption of multifractional behaviour when considered globally. Given the particular relevance of this issue to the work of the preceding chapters, this will be investigated further in the empirical exercise of Section 4.6.

4.3 Testing Methodology

The current section proposes several possible test statistics for determining whether a given sample of data is most consistent with a unifractal or a multifractal process. Section 4.3.1 summarises the basic testing problem and Section 4.3.2 presents the various test statistics proposed for distinguishing between unifractal and multifractal scaling behaviour.

4.3.1 A Summary of the Testing Problem

Whilst the previous informal graphical tests for unifractality versus multifractality are clearly flawed, the problems with this approach arise from the informal nature of the testing approach and the failure to account for the effects of estimation uncertainty, rather than the theoretical basis for the testing method. Indeed, the differences between $H(q)$ and $\tau(q)$ for unifractal and multifractal processes provide arguably the most logical method for distinguishing between the two classes of process and so will also be used

as a basis for the more formal testing procedure proposed here. The problem therefore becomes a question of formally testing whether the observed deviation from unifractal behaviour (either in terms of estimated $\tau(q)$ or estimated $H(q)$) is sufficiently large that it cannot be attributed solely to the inherent estimation uncertainty for the chosen estimator.

Clearly an assessment of the concavity of the scaling function or the dependence of the generalised Hurst exponent on q must be based on estimates of $\tau(q)$ or $H(q)$ for some set of values of q and cannot be made using the values of the functions at just a single value of q . This set of values will be denoted by the k -dimensional vector $\mathbf{q} = \{q_1, \dots, q_k\}$. In general the only condition that must be satisfied by this set of values is that $\mathbf{q} \in \mathbb{R}^k$. This implies that $H(q)$ and $\tau(q)$ can be estimated for both non-integer and non-positive values of q , although for such values the intuition becomes somewhat less clear for the moment scaling relationship of multifractal processes discussed in Section 4.2.1. For the specific test statistics proposed here, the values of the k elements of $\mathbf{q}$ must also be equally spaced, but more generally there is no need for this additional condition to hold.

The MF-DFA and MF-CMA estimators of Section 4.2.2 provide an estimate of $H(q)$ and $\tau(q)$ for a specific value of q , but via repeated application of the estimator for each of the k values in $\mathbf{q}$, the sets of estimated points $\{\hat{H}(q_1), \dots, \hat{H}(q_k)\}$ and $\{\hat{\tau}(q_1), \dots, \hat{\tau}(q_k)\}$ on the generalised Hurst exponent and scaling function are obtained. We therefore wish to construct test statistics that can be calculated from these sets of points in order to test whether the sample of data is most consistent with a unifractal or multifractal data generating process.

4.3.2 Hypothesis Tests of Unifractality Versus Multifractality

If viewing the testing problem from a traditional econometric modelling perspective, one way to construct a suitable test statistic may appear to be fitting a parametric functional form to the set of estimated points on $H(q)$ or $\tau(q)$ that allows for multifractal scaling

behaviour whilst nesting unifractal scaling as a special case for specific parameter values, with a test then based on the statistical significance of the coefficient(s) that allow for multifractal behaviour.

In practice however this approach is problematic, due to the need to fit a functional form to the generalised Hurst exponent or the scaling function. Most notably, there is the issue of choosing the most appropriate functional form under the alternative of multifractality; unlike the simpler class of unifractal processes, in the multifractal case the functional forms of $H(q)$ and $\tau(q)$ will differ from one process to another. If we are willing to assume a specific multifractal DGP under the alternative with a known parametric functional form then this issue can be solved, but then the test becomes one of unifractality versus a specific multifractal specification rather than a general test of unifractality versus multifractality.

Instead, this difficulty can be avoided by constructing test statistics based on the derivatives of the generalised Hurst exponent or scaling function and directly testing whether the gradient of $H(q)$ is zero (via the first derivative), or whether $\tau(q)$ is linear (via the second derivative). Without imposing specific functional forms on $H(q)$ or $\tau(q)$ it is not possible to obtain closed form expressions for their derivatives, however they can be approximated nonparametrically using numerical differentiation methods. This approach allows the method to provide a general test of unifractality versus multifractality, without the need to assume a specific multifractal process under the alternative.

Test statistics based on both the generalised Hurst exponent and the scaling function should in principle be equivalent given the identity $\tau(q) \equiv qH(q) - 1$ that relates the two functions. However, test statistics based on both functions have been included to demonstrate how tests statistics can be formed in each case and to explore in the Monte Carlo exercise of Section 4.5 whether there are any advantages in practice of basing tests on a specific representation of scaling behaviour.

Beginning with tests based on the generalised Hurst exponent, for a unifractal process

$H(q)$ is constant and so the first derivative, denoted $H'(q)$, should satisfy $H'(q) = 0 \forall q \in \mathbb{R}$. Numerous alternative numerical differentiation methods exist for approximating the first derivative, of which the central difference approach will be employed; this typically provides a more accurate approximation of the true derivative than the forward and backward difference approaches, but without being more computationally demanding when the function to be differentiated is of a single variable (as is the case with $H(q)$ in the current context).

Assume that $H(q)$ is estimated for all values of q in the vector $\mathbf{q} = \{q_1, \dots, q_k\}$, which consists of k equally spaced and strictly increasing values such that $q_i = q_1 + (i - 1)\Delta q$, where $\Delta q = q_2 - q_1$. The corresponding estimate of the generalised Hurst exponent for each $q_i \in \mathbf{q}$ are denoted by $\hat{H}(q_i)$. Then, using the central difference approach, the first derivative of $H(q)$ can be estimated numerically from the estimated generalised Hurst exponent values via :

$$\hat{H}'(q_i, \Delta q) = \frac{\hat{H}(q_{i+1}) - \hat{H}(q_{i-1})}{2\Delta q} \quad \text{for } i = 2, \dots, k - 1$$

where $\hat{H}'(q_i, \Delta q)$ denotes the approximate derivative of $H(q)$ at q_i , with an interval of length Δq between each of the points. Note that the use of the central difference method does not allow the derivative to be estimated at the end points of q_1 or q_k . A test statistic for unifractal distributional scaling can then be based on some function of these estimated derivatives $\hat{H}'(q_i, \Delta q)$ calculated over the $k - 2$ values.

Given the conditions $H'(q) = 0 \forall q \in \mathbb{R}$ for any unifractal process, there are many possible functions of $\hat{H}'(q_i, \Delta q)$ that could be used as statistics for testing the null hypothesis of unifractality. One possibility is given by the infimum of $\hat{H}'(q_i, \Delta q)$ taken over all $q \in \mathbf{q}$:

$$dH_{\inf} = \inf_{i=2,\dots,k-1} \left[\hat{H}'(q_i, \Delta q) \right] \quad (4.3.1)$$

For a true unifractal process, the population value of the $dH_{\inf}$ statistic in equation (4.3.1) equals zero, whereas for a multifractal process it is strictly negative, since $H(q)$ is a decreasing function of q . A test of the null hypothesis of unifractal scaling behaviour against the alternative of multifractal scaling can then be performed by testing $H_0 : dH_{\inf} = 0$ vs $H_1 : dH_{\inf} < 0$. In principle an alternative but closely related test statistic could be obtained by replacing the infimum of $\hat{H}'(q_i, \Delta q)$ in equation (4.3.1) with the supremum to give:

$$dH_{\sup} = \sup_{i=2,\dots,k-1} \left[\hat{H}'(q_i, \Delta q) \right]$$

where the null of unifractality again corresponds to $H_0 : dH_{\sup} = 0$, which is tested against the alternative $H_1 : dH_{\inf} < 0$.

For a unifractal process both the supremum and infimum of $\hat{H}'(q_i, \Delta q)$ must equal zero and so the two null hypotheses of $H_0 : dH_{\inf} = 0$ and $H_0 : dH_{\sup} = 0$ both correspond to unifractality. However, the alternative hypotheses of these two tests are subtly different; the infimum approach leads to rejection of the null of unifractality if the estimated generalised Hurst exponent exhibits multifractal scaling (i.e. $\hat{H}'(q_i, \Delta q) < 0$) for *some* $q \in \mathbf{q}$, whereas the supremum statistic leads to rejection if the estimate of $H(q)$ exhibits multifractal scaling *for all* $q \in \mathbf{q}$. Because of this distinction, the alternative of $H_1 : dH_{\sup} < 0$ corresponds specifically to multifractality, whereas the alternative of $H_1 : dH_{\inf} < 0$ corresponds to no unifractality, which includes multifractality as a possibility but does not guarantee it.

Whilst this distinction may prove advantageous in situations where the true data

generating process may be neither unifractal nor multifractal, if it is known (or can be assumed) that the true process is either unifractal or multifractal, then the dH_{inf} is likely to perform better in practice. To see why this is so, note that the value of the supremum-based statistic is determined by the smallest deviation from unifractal scaling, whereas the value of the infimum-based statistic is determined by the largest deviation from unifractal scaling⁹.

The estimates of $H(q)$ obtained from purely unifractal processes will still typically exhibit small deviations consistent with multifractal scaling due to the estimation error encountered when estimating $H(q)$. As a result, these small deviations from unifractal scaling are less informative about the true scaling properties of the data than large deviations. This was confirmed in practice by initially including both the dH_{inf} and dH_{sup} statistics in the Monte Carlo exercise of Section 4.5 and comparing their performance; in larger samples the differences in performance between tests based on the two statistics were found to be minor, but in smaller samples where the estimation error for $H(q)$ becomes larger, the infimum-based statistic was found to produce tests with superior empirical size and power properties.

Based on the above arguments, the dH_{inf} statistic was employed for the final Monte Carlo exercise of Section 4.5 and the alternative supremum-based statistic was omitted. Following the previous comments concerning the form of the alternative hypothesis for the dH_{inf} statistic, if it cannot be assumed a priori that the true data generating process is either unifractal or multifractal, then an additional check can be included to ensure that the data of interest appear to be consistent with either a unifractal or a multifractal process. This initial check could simply be performed using the same informal graphical methods previously employed in the literature; if the sample estimate of the generalised Hurst exponent $H(q)$ is non-increasing, then the data of interest are consistent with either a unifractal or a multifractal process. Having eliminated the possibility of processes

⁹Note however that in both cases rejection only occurs for negative sample values, so that only deviations consistent with a multifractal process are considered.

that are neither unifractal nor multifractal, rejection of $H_0 : dH_{\text{inf}} = 0$ now provides evidence directly in favour of multifractality.

Instead of basing a test statistic on the largest deviation from unifractal scaling exhibited by $H(q)$ over the chosen domain of q , a test statistic could also be constructed based on the average deviation over all $q \in \mathbf{q}$. Such a statistic can be calculated as:

$$dH_{\text{avg}} = \frac{1}{k-2} \sum_{i=2}^{k-1} \left| \hat{H}'(q_i, \Delta q) \right| \quad (4.3.2)$$

This alternative dH_{avg} statistic measures the average deviation in absolute value from unifractal scaling of $\hat{H}(q)$ over the chosen domain of q and again, for a unifractal process the value of the dH_{avg} test statistic in equation (4.3.2) should be equal to zero. Note that for the dH_{avg} statistic it is necessary to take the absolute value of each $\hat{H}'(q_i, \Delta q)$ before calculating the average, to ensure that positive and negative values from a non-unifractal process do not cancel each other out and produce a value of dH_{avg} close to zero. This implies that for a multifractal process dH_{avg} will be strictly positive and so a test of the null hypothesis of unifractal scaling behaviour against the alternative of multifractal scaling can then be performed by testing $H_0 : dH_{\text{avg}} = 0$ vs $H_1 : dH_{\text{avg}} > 0$. As with the dH_{inf} statistic above, the alternative of $H_1 : dH_{\text{avg}} > 0$ technically corresponds to no unifractality, rather than multifractality. A test for unifractality versus multifractality can however be implemented as discussed above, by first checking that the data of interest appear to be consistent with either a unifractal or multifractal process.

A closely related statistic to the dH_{avg} statistic above could be constructed based on the average squared deviation from unifractal scaling over the chosen domain of q , rather than the average absolute deviation, with the resulting statistic given by:

$$dH_{\text{ls}} = \frac{1}{k-2} \sum_{i=2}^{k-1} \hat{H}'(q_i, \Delta q)^2$$

The basic intuition behind this form of test statistic is identical to that for the previous dH_{avg} statistic, with the key difference between the two being that the dH_{ls} statistic places more weight on large deviations from unifractal scaling relative to the dH_{avg} statistic, and relatively less weight on smaller deviations.

In light of the earlier discussion concerning the relative informativeness of large and small deviations from unifractal scaling, it might be expected that the dH_{ls} statistic would perform better than the dH_{avg} average statistic due to the higher weight attached to large deviations, which are likely to be more informative than smaller deviations. In practice however the dH_{ls} statistic performs almost identically to the dH_{avg} statistic in Monte Carlo exercises using the methodology described in Section 4.5.1. Therefore, in order to keep the following discussion concise, attention is restricted to the infimum and average type statistics presented above for the remainder of the work.

For tests based on the scaling function a measure of concavity is required, which can be provided by the second derivative, $\tau''(q)$. The linear scaling function of a unifractal process implies that $\tau''(q) = 0 \forall q \in \mathbb{R}$. When combined with the property that $\tau(q)$ is non-decreasing for both unifractal and multifractal processes, the strictly concave scaling function for a multifractal process must satisfy $\tau''(q) < 0 \forall q \in \mathbb{R}^1$.

It is assumed that $\tau(q)$ is estimated at the same vector of k equally spaced values $\mathbf{q} = \{q_1, \dots, q_k\}$, with the resulting estimates denoted by $\hat{\tau}(q_i)$ for $i = 1, \dots, k$. An estimate of the second derivative of $\tau(q)$ at $q = q_i$ can then be obtained by applying numerical differentiation methods to the estimates $\hat{\tau}(q_i)$, via:

$$\hat{\tau}''(q_i, \Delta q) = \frac{\hat{\tau}(q_{i+1}) - 2\hat{\tau}(q_i) + \hat{\tau}(q_{i-1}))}{(\Delta q)^2} \quad \text{for } i = 2, \dots, k-1$$

Whilst testing approaches based on $\tau(q)$ and $H(q)$ should be equivalent in principle,

the use of second order numerical differentiation to estimate $\tau''(q)$ may be expected to produce poorer approximations to the true second derivative than those obtained for the first derivative of $H(q)$ employed by the statistics above. However, the differences in test statistic performance between these two approaches do not appear to be large in practice and statistics based on $\tau(q)$ are more consistent with the previous graphical testing methods employed in the literature (which are typically based on $\tau(q)$ and not $H(q)$), thus justifying their inclusion in the analysis.

As with the previous test statistics based on $H'(q)$, various functions of $\hat{\tau}''(q_i, \Delta q)$ could be employed as test statistics in the current context. However, attention is restricted to two examples constructed from $\hat{\tau}''(q_i, \Delta q)$ in an analogous way to those based on $H'(q_i, \Delta q)$ in equation (4.3.2):

$$d\tau_{\inf} = \inf_{i=2, \dots, k-1} [\hat{\tau}''(q_i, \Delta q)] \quad \text{or} \quad d\tau_{\text{avg}} = \frac{1}{k-2} \sum_{i=2}^{k-1} |\hat{\tau}''(q_i, \Delta q)| \quad (4.3.3)$$

The interpretation of the $d\tau_{\inf}$ and $d\tau_{\text{avg}}$ test statistics above is identical to that of the earlier test statistics in equation (4.3.2); the $d\tau_{\inf}$ statistic measures the largest deviation from unifractal scaling behaviour over the chosen domain of q , whereas the $d\tau_{\text{avg}}$ statistic measures the average absolute deviation from unifractal scaling over the same domain. As with the previous dH_{avg} statistic, it is necessary to use the absolute values of $\hat{\tau}''(q_i, \Delta q)$ in the construction of the $d\tau_{\text{avg}}$ statistic, to ensure that a non-linear scaling function with $\hat{\tau}''(q_i, \Delta q) < 0$ for some q and $\hat{\tau}''(q_i, \Delta q) > 0$ for other q cannot be incorrectly classified as unifractal.

The linear scaling function of a unifractal process implies that the true population values of both the $d\tau_{\inf}$ and $d\tau_{\text{avg}}$ test statistics should be equal to zero, whereas the strict concavity of the scaling function for a multifractal process implies $d\tau_{\inf} < 0$ and $d\tau_{\text{avg}} >$

0. A test of the null hypothesis of unifractality against the alternative of multifractality¹⁰ can then be performed by testing $H_0 : d\tau_{\text{inf}} = 0$ vs $H_1 : d\tau_{\text{inf}} < 0$ or $H_0 : d\tau_{\text{avg}} = 0$ vs $H_1 : d\tau_{\text{avg}} > 0$.

4.4 Resampling Methods for Obtaining Test Statistic Distributions

In order to obtain critical values or p-values for the proposed test statistics, the distribution of the statistics under the null hypothesis of unifractality must be obtained. Unfortunately in the current context this is complicated by several factors discussed in Section 4.4.1, which any suitable resampling method must take into account. The model-based bootstrap resampling approach proposed to overcome these issues is discussed in detail in Section 4.4.2, followed finally by a brief discussion of more conventional resampling schemes in Section 4.4.3.

4.4.1 Issues Affecting the Choice of Resampling Method

The first is that the distributions of the test statistics proposed in Section 3 are non-standard and their rates of convergence are not known. This necessitates the use of nonparametric distributions to approximate the finite sample distributions of the test statistics, which can be obtained using appropriate resampling methods.

Secondly, except in special cases for specific parameter values, both unifractal and multifractal processes will exhibit serial dependence; this serial dependence frequently persists at long horizons and can be linear in the case of unifractal processes and both linear or non-linear for multifractal processes. Therefore, any resampling method employed in the current context to obtain distributions for the test statistics must be able

¹⁰As with the statistics based on $H(q)$ above, the alternative hypotheses of $d\tau_{\text{inf}} < 0$ and $d\tau_{\text{avg}} > 0$ technically correspond to no unifractality rather than multifractality and should be combined with the same type of additional check described above.

to reproduce this dependence structure. Although the serial dependence structure is not of direct interest, this issue is particularly critical in the current testing environment, because the serial dependence in unifractal and multifractal processes is intimately linked with their scaling properties, which provide the theoretical basis of the proposed test statistics. As a result, any resampling method that cannot reproduce the serial dependence structure will also be unable to reproduce the scaling properties and is likely to lead to invalid inference.

Finally, the value of any of the proposed test statistics calculated from the estimates $\hat{H}(q)$ or $\hat{\tau}(q)$ will depend not only on the sample of data from which it is calculated, but also on the specific method used to estimate these functions from that sample. From the preceding discussion of Sections 4.2 and 4.3, it should be clear in the present context that this includes the choice of estimator itself (such as MF-DFA or MF-CMA) and also additional choices including the set of values for q over which the functions are estimated and the minimum and maximum window sizes (denoted s_{min} and s_{max} in Section 4.2) used when implementing the estimators.

This distinction between the estimator itself and the estimation *method* as a whole has some parallels with the work of Giacomini and White (2006) on tests of conditional predictive ability; there, the authors make a distinction between forecasting *models* and forecasting *methods*, with the latter including all additional choices that must be made in order to obtain a forecast from the chosen model, including the estimation procedure selected and length of the estimation window employed. In the testing environment considered there, all these aspects of the forecasting method will affect the asymptotic distribution of the test and the same is true for the estimation method in the current testing environment.

In practice, these factors are too numerous and their effects on the test statistic distributions too complex for pivotal versions of the test statistics to be derived that are independent of these factors. This precludes the tabulation of critical values and the

distributions obtained for the test statistics must be conditional on these aspects of the estimation method. Applying equivalent terminology to that of Giacomini and White (2006) to the current context, it could be said that the proposed testing methodology must account for the effects of the estimation method for $H(q)$ or $\tau(q)$ when deriving the null distribution of the test statistics. Furthermore, because the effects of the estimation method are accounted for within the testing procedure, it is possible to directly compare the performance of the tests for different choices of estimation method, as done in the Monte Carlo exercise of Section 4.5.

4.4.2 Model-based Bootstrap Resampling

In response to the requirements discussed above, a resampling approach was developed for obtaining the null distributions of the test statistics that can be considered as a model-based bootstrap approach, combining aspects of both Monte Carlo and bootstrap methods. Whilst this resampling approach potentially has more general applicability in other testing situations, in order to motivate the use of this method in the current context some important characteristics specific to the current testing environment must be emphasised.

Firstly, all of the proposed test statistics are based solely on the scaling properties of the process, which are completely described by either the generalised Hurst exponent, $H(q)$, or equivalently the scaling function, $\tau(q)$. As such, none of the test statistics depend directly on any other characteristics of the underlying process. Secondly, the null hypotheses for every test statistic proposed in Section 4.3.2 corresponds to that of a unifractal process, which is always tested against an alternative of either multifractal or non-unifractal scaling. From Theorem 4.1, every unifractal process has an identical functional form for the generalised Hurst exponent and scaling function, parameterised entirely by the simple Hurst exponent, H . Finally, under the null of unifractal scaling, the true value of all proposed test statistics is equal to zero.

Whilst the true scaling properties are identical for any unifractal process with the same simple Hurst exponent, the functions $H(q)$ and $\tau(q)$ must be estimated in practice for a given sample of data. Due to the effects of estimation error, the sample values of these functions obtained from multiple realisations of the same process will differ, potentially leading to the spurious multifractality previously discussed in Section 4.2.3. We therefore wish to model this estimation uncertainty around the true values of $H(q)$ and $\tau(q)$ by obtaining the distribution of the relevant test statistics under the null of unifractality using an appropriate resampling method. This will allow us to determine whether the observed deviation from unifractal scaling is sufficiently large to be interpreted as evidence of multifractal scaling, or whether it can be explained simply by estimation error for a process that is in fact purely unifractal.

The discussion above suggests that we can begin by estimating the simple Hurst exponent from the original sample of data, under the assumption that the data are consistent with the null hypothesis of unifractality¹¹. Next, a large number of independent sample paths are simulated from a specific theoretical unifractal process, with a simple Hurst exponent equal to the estimated value of H . For each of these simulated sample paths, the value of the relevant test statistic can be calculated and the empirical distribution of these test statistic values over the set of simulated series can be obtained. This empirical distribution can then be used as an estimate of the distribution of the test statistic of interest under the null of unifractality. Formally, the steps of the resampling method are as follows:

1. The simple Hurst exponent, H , is estimated from the original sample of data, together with an additional parameter vector θ , containing any other parameters required to fully characterise the chosen unifractal process used for simulation in stage 2. The resulting estimates are denoted by $\hat{H}$ and $\hat{\theta}$ respectively.

¹¹At this stage it is not necessary to assume a specific process a priori, since the various estimators available for the simple Hurst exponent are valid for any unifractal process.

2. The estimated parameter values $\hat{H}$ and $\hat{\theta}$ are used to generate R independent realisations from the chosen unifractal data generating process, with each of the simulated series having the same length as the original time series of data.
3. For each of the R independent unifractal realisations the functions $H(q)$ or $\tau(q)$ are estimated, with the resulting estimates denoted by $\tilde{H}(q)_1, \dots, \tilde{H}(q)_R$ and $\tilde{\tau}(q)_1, \dots, \tilde{\tau}(q)_R$. From these estimates, the sample value of the chosen test statistic from Section 3 is then calculated from each of these R estimates, with these test statistic sample values for the chosen test statistic denoted generically by $\tilde{T}_1, \dots, \tilde{T}_R$ for notational simplicity, rather than for example $\widetilde{dH}_{\text{inf},1}, \dots, \widetilde{dH}_{\text{inf},R}$.
4. Sample estimates of $H(q)$ and $\tau(q)$ are obtained from the original sample of data using an *identical* estimation method to that employed in the previous step to obtain the estimates for the simulated realisations. The sample estimates are denoted $\hat{H}(q)$ and $\hat{\tau}(q)$ and are used to calculate the sample value of the chosen test statistic, denoted by $\hat{T}$.
5. The empirical cumulative distribution function of the R bootstrap test statistic values $\tilde{T}_1, \dots, \tilde{T}_R$ denoted by $\tilde{F}(T)$ is calculated as:

$$\tilde{F}(T) = \frac{1}{R} \sum_{i=1}^R I(\tilde{T}_i \leq T)$$

where $I(\tilde{T}_i \leq T)$ is the indicator function taking the value of 1 if $\tilde{T}_i \leq T$ and 0 otherwise. The empirical distribution $\tilde{F}(T)$ is then used as an estimate of the distribution of the test statistic under the null of unifractal scaling behaviour.

6. The two infimum-based statistics are one-sided tests with rejection regions on the left side of the distribution and so the p-value of the sample value $\hat{T}$ can be calculated as $\tilde{F}(\hat{T})$, with the null of unifractality being rejected if the sample p-value is less than the chosen nominal significance level. The average-based statistics are

again one-sided, but with rejection regions on the right side of the distribution. Sample p-values can then be calculated as $1 - \tilde{F}(\hat{T})$, with the null rejected again if the sample p-value is less than the chosen nominal significance level. Alternatively, in both cases the empirical inverse cumulative distribution function can be calculated and used to obtain critical values.

Given that the original aim was to develop a general test for unifractality versus multifractality, the need to choose a specific unifractal data generating process for the resampling scheme may seem restrictive, however this can be justified in the following way. From the above discussion it is clear that all proposed test statistics are solely functions of either $H(q)$ or $\tau(q)$ over some discrete set of values for q . For any unifractal process under the null hypothesis with the same true value for the simple Hurst exponent, not only are the population values of all proposed test statistics identical, but so too are the true functional forms of $H(q)$ or $\tau(q)$ for any choice of q .

Of course in practice we do not observe the true values of the generalised Hurst exponent and scaling function, but obtain estimates of these functions; these estimates can be considered as consisting of the true value (identical for all unifractal processes with a common value of H) plus the estimation error around this true value, the latter of which is what the resampling method aims to model. However, if we impose the assumption that the form of the estimation error is identical for all unifractal processes with a common simple Hurst exponent, then it is irrelevant which specific unifractal data generating process is used to obtain critical values for the test statistics.

Crucially it should be noted that this assumption of identical estimation error across different unifractal processes is made conditional on the complete estimation method employed and the value of the simple Hurst exponent. As a result, it still permits the estimation error to depend on the value of H , the sample size, the values of q for which $H(q)$ or $\tau(q)$ are estimated, the choice of estimator employed and all other aspects of the estimation method previously discussed at the beginning of the current

section. It should also be noted that the alternative hypothesis remains composite in that it is not necessary to specify a particular multifractal or non-unifractal process to test against, except when testing the empirical power of the testing method against a specific alternative in a Monte Carlo exercise such as that in Section 5.

Finally, it is critical that both the sample size and the method used to estimate $H(q)$ and $\tau(q)$ for each of the R generated unifractal series in stage 3 above is identical to that used to obtain the sample estimate of $H(q)$ or $\tau(q)$ from the original time series of data in stage 4. As previously discussed, the null distributions of the test statistics depend on the complete estimation method employed for $H(q)$ and so using different methods at these two stages will result in a null distribution that does not correspond to the distribution of interest.

For the Monte Carlo exercises of Section 4.5, the specific unifractal process used in stage 3 of the resampling algorithm to obtain the null distributions of the test statistics was the fractional Gaussian noise (fGn). Whilst alternative unifractal processes could be employed, the fGn and its corresponding cumulative sum series, the fractional Brownian motion (fBm), are by far the most commonly used unifractal process in the finance literature, and indeed elsewhere.

The fractional Gaussian noise is the increment series of the fractional Brownian motion, which can be viewed as a generalisation of the standard Brownian motion allowing for long-range dependence in the values of the fGn increment series. More formally, an fGn process $\{X_t : t = 1, \dots, T\}$ is a series of Normal random variables with mean 0 and autocovariance function:

$$E[X_{t+k}X_t] = \gamma(k) = \frac{\sigma^2}{2} \{|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}\}$$

where H is the simple Hurst exponent of the process¹². For the special case for $H = 0.5$ the fGn has independent increments and simplifies to the standard Gaussian noise, with the corresponding fBm simplifying to the standard Brownian motion. Numerous algorithms are available for the simulation of sample paths from a fGn/fBm process, with the circulant embedding method used here, which was first properly formalised by Dietrich and Newsam (1997); this algorithm is applicable more generally for the simulation of any stationary Gaussian process and generates exact rather than approximate sample paths, but is substantially faster than other exact simulation methods such as the Cholesky decomposition approach.

For the fGn process, the only additional parameter required to completely characterise the process is the standard deviation, σ , which can be easily estimated from the original data using an appropriate heteroskedasticity and autocorrelation consistent estimator to allow for long-range dependence in the series. In practice however the estimates of $H(q)$ or $\tau(q)$ obtained from the chosen MF-DFA and MF-CMA estimators should not be affected by the unconditional variance of the series and so the choice of unconditional variance used when simulating the fGn processes is not critical. Sample estimates of the simple Hurst exponent H for a unifractal process can be obtained using the unifractal analogues of the MF-DFA or MF-CMA estimators of Section 4.2.2, which as discussed earlier are obtained from the corresponding multifractal estimators for the special case of $q = 2$.

4.4.3 Alternative Resampling Approaches

Instead of the model-based parametric bootstrap resampling approach proposed above, it would in principle be possible to employ alternative resampling schemes to obtain

¹²For the fGn the same parameter controls both the scaling and long-range dependence properties of the process, which is a property of the class of ‘self-affine’ processes, of which the fBm and fGn are members. More generally, unifractal processes may in principle possess independent scaling and dependence properties. However, the construction of such generalised processes is more complex and so few examples exist in the literature, with a notable example being that of Gneiting and Schlather (2004).

the null distribution of the test statistic(s). In particular, nonparametric resampling methods such as block bootstrap or sub-sampling could be employed. As argued above, such a nonparametric resampling approach is not required in the current context due to the specific characteristics of the testing problem. However, it is worth noting that the application of such nonparametric resampling methods in the current context is not just unnecessary, but also extremely problematic in practice.

The key problem encountered is the serial dependence structure of unifractal and multifractal processes; the class of unifractal processes considered under the null will display long-range linear dependence, except in the specific case where the simple Hurst exponent satisfies $H = 0.5$, for which the process is serially independent. Multifractal processes display more complex serial dependence, with the most applicable examples in the context of finance (such as the MRW and Markov-switching multifractal of Calvet and Fisher, 2004) exhibiting long-range dependence in the variance, even if the process exhibits no linear serial dependence. As a result, any resampling scheme employed to estimate the distribution of the test statistics under the null must allow for and preserve such patterns of serial dependence in the data.

This proves especially problematic for two reasons; the first is simply that the length of this linear or non-linear serial dependence is often very long. Secondly, although the functions $H(q)$ and $\tau(q)$ are directly concerned with how the moments of the process scale across different sampling intervals, this is indirectly but strongly linked to the dependence structure of the process. As a result, even relatively small changes in the original dependence structure of the process caused by inappropriate resampling methods will have a potentially severe effect on the resulting estimates of the scaling properties. The scale of this problem is such that the use of standard block bootstrap resampling approaches to estimate the distribution of the test statistics was found to perform very poorly in Monte Carlo exercises, even for sample sizes much larger than those for which the model-based bootstrap approach performed well.

4.5 Monte Carlo Exercises

The current section uses simulated unifractal and multifractal processes with known properties to examine the performance of the proposed testing methodology in terms of empirical size and power in a variety of situations. Section 4.5.1 begins with a summary of the Monte Carlo testing methodology employed throughout the current section to estimate the empirical size and power of the tests, including the default choices for the estimation method used for $H(q)$ and $\tau(q)$. Section 4.5.2 reports and discusses the size and power properties of the testing methodology using these default choices for various sample sizes. Sections 4.5.3 and 4.5.4 then examine the effect of changing these default choices on the performance of the tests, with the first considering changes in the estimation method employed for the generalised Hurst exponent or scaling function and the second examining the empirical size and power of the tests with a wider range of unifractal and multifractal processes under the null and alternative respectively.

4.5.1 Monte Carlo Methodology

A standard Monte Carlo approach is employed for obtaining estimates of the empirical size and power of the various test statistics under different conditions. When calculating empirical size under the null of unifractality, the testing methodology must be applied to realisations from some unifractal data generating process consistent with the null hypothesis. For this purpose the fractional Gaussian noise previously discussed in Section 4.4 is employed. The model-based bootstrap resampling scheme is applied to the generated data, with $R = 1,000$ resamples used in stages 2 and 3 of the algorithm to estimate the distributions of the chosen test statistics under the null of unifractality. From these empirical null distributions, the outcome of each test at several nominal significance is recorded. This process is repeated for M Monte Carlo repetitions in total, thus producing M test outcomes for each nominal significance level. The empirical size

of the test is then calculated as the percentage of rejections at each significance level over these M realisations.

When calculating empirical power under the alternative of multifractality the same procedure is employed, but with the unifractal fGn process replaced by a multifractal process, with the default choice being the log-normal variant of the multifractal random walk (MRW) previously mentioned in Section 4.2. Following Bacry et al. (2001), a discrete time formulation of the MRW $X_{\Delta t, t}$ with time discretisation step Δt ¹³ can be used to simulate sample paths from the MRW process. This discrete time formulation can be constructed by summing the series of $t/\Delta t$ random variables:

$$X_{\Delta t, t} = \sum_{k=1}^{t/\Delta t} \delta X_{\Delta t, k}$$

The $\delta X_{\Delta t, k}$ are the corresponding increment series and can be expressed as:

$$\delta X_{\Delta t, k} = \epsilon_{\Delta t, k} e^{\omega_{\Delta t, k}}$$

where $\epsilon_{\Delta t, k}$ is a Gaussian noise with variance $\sigma^2 \Delta t$ that is independent of $\omega_{\Delta t, k}$. The $\omega_{\Delta t, k}$ term can be viewed as the logarithm of the stochastic volatility of the increment process $\delta X_{\Delta t, k}$ and is a stationary Gaussian process with autocovariance function of the form:

$$Cov(\omega_{\Delta t, i}, \omega_{\Delta t, j}) = \lambda^2 \ln \rho_{\Delta t} [|i - j|]$$

¹³The time discretisation step could be set to value of $\Delta t = 1$ without loss of generality by simply normalising the units of time used for measurement.

where

$$\rho_{\Delta t} [|i - j|] = \begin{cases} \frac{T}{(|i-j|+1)\Delta t} & \text{if } |i - j| \leq T/\Delta t - 1 \\ 1 & \text{otherwise} \end{cases}$$

where T is the integral time, beyond which $Cov(\omega_{\Delta t,i}, \omega_{\Delta t,j}) = 0$. Finally, the mean of the process $\omega_{\Delta t,k}$ must be chosen to satisfy $E[\omega_{\Delta t,k}] = -\lambda^2 \ln(T/\Delta t)$.

The MRW was selected due to its suitability for financial applications (see Muzy et al., 2001) and the fact that the strength of multifractality is controlled directly by the single parameter λ^2 , thus permitting the relationship between power and the strength of multifractality to be easily studied. The alternative Markov-switching multifractal (MSM) model of Calvet and Fisher (2004) and Calvet et al. (2006) is also employed in Section 4.5.4 to confirm that the testing procedure has power against a second multifractal process commonly used in financial applications.

For both the unifractal fGn and MRW processes the default parameter values were chosen based on parameter estimates obtained from daily return series for a range of financial assets, however the robustness of the results to different choices for these parameter values is explored in Section 4.5.4. For the fGn process the default values used were $H = 0.5$ for the simple Hurst exponent and $\sigma = 0.1$ for the standard deviation. The value of $H = 0.5$ implies linear independence and was chosen based on the finding that estimates of H for financial returns are typically close to 0.5, reflecting the weak linear dependence normally found in the level of returns. For the MRW process the default parameter values were again $\sigma = 0.1$ for the standard deviation and $\lambda^2 = 0.025$ for the strength of multifractality parameter¹⁴. Again, the MRW realisations were constructed to possess linear independence, though by construction the process will display non-linear serial dependence, most notably in the variance.

¹⁴The additional parameters for the integral timescale, T , and sampling interval Δ were set to $T = 5000$ and $\Delta = 1$, though these are of less direct interest and are only defined relative to the sampling frequency of the data, which is arbitrary when dealing with simulated series.

The fact that the R bootstrap sample paths are generated randomly for each of the M original series every time the tests were performed introduces a very small amount of randomness in the reported size and power values. However, the chosen values of $R = 1,000$ bootstrap resamples and $M = 1,000$ Monte Carlo repetitions were found to be sufficient to produce empirical size and power values that are identical to the nearest percentage point on repeated runs of the Monte Carlo exercise.

As previously discussed, the degree of spurious multifractality due to the estimation error for $H(q)$ or $\tau(q)$ will be dependent on the estimation method employed, which is a combination of numerous factors rather than simply the choice of estimator itself. The first factor is the vector of values of q over which the functions are estimated, previous denoted by $\mathbf{q}$ in Section 4.3. Given that the vector of values must be equally spaced for the current derivative-based test statistics, $\mathbf{q}$ can be completely characterised by the minimum value, the maximum value and the grid or step size between each value, which are denoted by q_{min} , q_{max} and q_{grid} respectively.

The estimates of $H(q)$ obtained for values of q that are large in absolute terms are typically become very unreliable in finite samples, but at the same time it is necessary to consider the behaviour of $H(q)$ or $\tau(q)$ over a sufficient range of q to provide an accurate measure of the properties of these functions. Based on these considerations, the default choice was $q_{min} = 0$ and $q_{max} = 2.5$ with a grid size of $q_{grid} = 0.1$, which was found to produce constantly strong performance for the sample sizes considered here. Additional choices for q were also considered up to a maximum range of $0 \leq q \leq 5$ in Section 4.5.3¹⁵, to check the robustness of the testing methodology with alternative choices of $\mathbf{q}$. Larger sample sizes for the Monte Carlo exercises would permit more accurate estimation for larger values of q_{max} and allow the range of $\mathbf{q}$ to be expanded, however it was decided to focus on test performance with sample sizes that are somewhat small by the standards of the multifractal literature, in order to demonstrate the performance of the tests under

¹⁵Values of q_{max} in the range $4 \leq q_{max} \leq 5$ are common in empirical studies of multifractality in financial data, such as Di Matteo (2007), Muzy et al. (2001) or Bacry et al. (2008).

more challenging conditions. Negative values of q were also tested and also resulted in tests with good size and power properties, but performance was very similar to that when using strictly positive values for q and so these additional results are not reported to keep the discussion concise.

Of the two estimators discussed in Section 4.2.2, the MF-DFA estimator with simple linear detrending (sometimes referred to as the MF-DFA-1 estimator in the literature) is used as the default choice for the Monte Carlo exercises, since in most cases it was found to produce tests with marginally better empirical size and power properties than the alternative MF-CMA estimator. The differences between the two estimators are however often small and are discussed in detail in Section 4.5.3. The minimum and maximum scale or window sizes, previously denoted s_{min} and s_{max} , were set approximately according to the rules of thumb discussed in Section 4.2.2, with $s_{min} = 5$ in all cases and $s_{max} \approx N/15$, where N is the sample size.

4.5.2 Core Results & the Effects of Sample Size

Table 4.1 presents the empirical size and power of the proposed test statistics obtaining using the default choices detailed above with sample sizes of 5,000, 2,500 and 1,000. Whilst these appear to be very large sample sizes by normal econometric standards, it must be remembered that estimation of the generalised Hurst exponent and scaling function is based on the properties of the process for a range of timescales, ranging from that of the original data up to much longer sampling intervals. Therefore accurate estimation of the scaling properties of a sample of data requires a substantial number of observations, with minimum sample sizes in the low thousands normally suggested in the literature. As such, the need to use sample sizes of this magnitude should not be viewed as a limitation of the proposed testing methodology, but an inherent limitation in the study of this type of process.

From Table 4.1 it can be seen that for the larger sample sizes of $N = 5,000$ and

$N = 2,500$ the empirical size of the tests deviates very little from the specified nominal significance levels, regardless of the nominal level chosen. Furthermore, there seems to be little difference between the size properties of the tests at these two larger sample sizes, except arguably at the 1% level where the differences are more pronounced. At the smallest sample size of $N = 1,000$ the empirical sizes do move further from their nominal levels, but interestingly the deviations from the nominal levels appear smallest at the 1% level. It should again be remembered that a sample size of 1,000 is extremely small compared to those generally used in the literature on unifractal and multifractal processes.

Turning next to the empirical power of the tests, as expected power declines as the sample size decreases, ranging between 86% and 95% for $N = 5,000$ and dropping to between 53% and 73% for the smallest sample size of 1,000. It should also be noted that for the larger sample sizes the power of the tests appears to be much more consistent across the different nominal significance levels than for $N = 1,000$. However, this will also be influenced to some extent by the differences in empirical size, since the reported empirical power values are not size adjusted. It can be seen that particularly at the 10% and 5% levels the empirical size values for $N = 1,000$ are typically lower than the nominal levels, sometimes substantially so, which will in turn reduce the power of the tests in these cases relative to a test performed at the true nominal significance level.

For any of the given sample sizes the relative performance of the different test statistics is largely consistent, with only minor differences in empirical size and power between them. This suggests both that the approaches based on the generalised Hurst exponent $H(q)$ and the scaling function $\tau(q)$ are indeed equivalent in practice as well as in theory and that tests based on the average and infimum are approximately equal in terms of empirical size and power. The same was not true for the supremum-based statistics that were excluded from the Monte Carlo exercises for the reasons given in Section 4.3.2, which

Table 4.1: Empirical size and power for default testing methodology

Nominal sig. level	Empirical size			Empirical power		
	0.10	0.05	0.01	0.10	0.05	0.01
N = 5,000						
dH_{inf}	0.097	0.044	0.009	0.937	0.925	0.887
dH_{avg}	0.098	0.044	0.008	0.913	0.900	0.860
$d\tau_{\text{inf}}$	0.095	0.046	0.008	0.946	0.931	0.901
$d\tau_{\text{avg}}$	0.100	0.044	0.005	0.908	0.891	0.858
N = 2,500						
dH_{inf}	0.106	0.051	0.007	0.876	0.829	0.744
dH_{avg}	0.099	0.047	0.007	0.833	0.807	0.737
$d\tau_{\text{inf}}$	0.103	0.050	0.008	0.884	0.843	0.767
$d\tau_{\text{avg}}$	0.096	0.053	0.008	0.838	0.811	0.740
N = 1,000						
dH_{inf}	0.098	0.046	0.010	0.722	0.672	0.534
dH_{avg}	0.092	0.045	0.012	0.712	0.668	0.563
$d\tau_{\text{inf}}$	0.098	0.043	0.011	0.739	0.688	0.572
$d\tau_{\text{avg}}$	0.093	0.042	0.009	0.720	0.679	0.575

often produced empirical size values below the nominal significance level, particularly for the smallest sample size of 1,000.

4.5.3 Robustness to Changes in the Estimation Method

Given the importance of the estimation method for $H(q)$ and $\tau(q)$ in determining the estimation error and the degree of spurious multifractality it is crucial to check that the proposed testing procedure maintains good size and power properties for a range of reasonable estimation methods. The issue of sample size, which can be considered as part of the estimation method, has already been covered to some extent in the previous subsection. We now extend this robustness analysis by examining changes in the vector of values $\mathbf{q}$ over which the generalised Hurst exponent and scaling function are estimated and also exchanging the default MF-DFA estimator for the alternative MF-CMA estimator.

Table 4.2 contains empirical size and power values for various choices of the vector

Table 4.2: Empirical size and power for different choices of $\mathbf{q}$

Nominal sig. level	Empirical size			Empirical power		
	0.10	0.05	0.01	0.10	0.05	0.01
(a) For $q_{\min} = 0$ and $q_{\max} = 2.5$						
$q_{\text{grid}} = 0.05$						
dH_{inf}	0.098	0.044	0.008	0.940	0.926	0.889
dH_{avg}	0.097	0.047	0.005	0.914	0.902	0.862
$d\tau_{\text{inf}}$	0.098	0.045	0.008	0.950	0.934	0.902
$d\tau_{\text{avg}}$	0.102	0.050	0.004	0.911	0.892	0.860
$q_{\text{grid}} = 0.10$						
dH_{inf}	0.097	0.044	0.009	0.937	0.925	0.887
dH_{avg}	0.098	0.044	0.008	0.913	0.900	0.860
$d\tau_{\text{inf}}$	0.095	0.046	0.008	0.946	0.931	0.901
$d\tau_{\text{avg}}$	0.100	0.044	0.005	0.908	0.891	0.858
$q_{\text{grid}} = 0.25$						
dH_{inf}	0.091	0.044	0.006	0.933	0.919	0.883
dH_{avg}	0.096	0.045	0.005	0.908	0.894	0.855
$d\tau_{\text{inf}}$	0.090	0.043	0.005	0.938	0.922	0.890
$d\tau_{\text{avg}}$	0.105	0.050	0.005	0.901	0.887	0.855
(b) For $q_{\min} = 0$ and $q_{\max} = 5$						
$q_{\text{grid}} = 0.05$						
dH_{inf}	0.095	0.044	0.008	0.954	0.939	0.905
dH_{avg}	0.102	0.047	0.007	0.925	0.907	0.873
$d\tau_{\text{inf}}$	0.093	0.044	0.008	0.970	0.958	0.927
$d\tau_{\text{avg}}$	0.096	0.048	0.007	0.938	0.920	0.889
$q_{\text{grid}} = 0.10$						
dH_{inf}	0.097	0.044	0.009	0.952	0.940	0.908
dH_{avg}	0.104	0.046	0.005	0.923	0.905	0.872
$d\tau_{\text{inf}}$	0.095	0.046	0.008	0.972	0.962	0.932
$d\tau_{\text{avg}}$	0.094	0.049	0.009	0.937	0.917	0.886
$q_{\text{grid}} = 0.25$						
dH_{inf}	0.093	0.045	0.008	0.949	0.936	0.909
dH_{avg}	0.097	0.048	0.005	0.918	0.902	0.872
$d\tau_{\text{inf}}$	0.093	0.048	0.008	0.971	0.958	0.930
$d\tau_{\text{avg}}$	0.094	0.048	0.010	0.933	0.913	0.875

$\mathbf{q}$. Results are only presented for the sample size of 5,000 in order to conserve space, but the same analysis was also performed for the previous sample sizes of 2,500 and 1,000,

with similar results obtained. Table 4.2a reports results for the same default values of $q_{min} = 0$ and $q_{max} = 2.5$, but varying the grid size q_{grid} used from the default value of 0.1 to include the two additional values of 0.05 and 0.25. From the reported values it is clear that varying the grid size within the chosen range has a minimal effect on the performance of the test, with performance remaining consistent across the range $0.05 \leq q_{grid} \leq 0.25$.

Table 4.2b employs the same range of grid sizes as Table 4.2a, but increasing the maximum value of q to $q_{max} = 5$. Again, the performance of the tests remains consistent both across the different grid sizes used in Table 4.2b for the value of $q_{max} = 2.5$ and also when comparing these results with those previously reported in Table 4.2a for the default value of $q_{max} = 2.5$. Whilst more extreme choices for the vector $\mathbf{q}$ (particularly using large values of q_{max}) are likely to have an adverse effect on the performance of the proposed testing methodology, the good size and power properties of the tests do seem to be preserved across a variety of reasonable choices for $\mathbf{q}$.

Table 4.3 compares the empirical size and power obtained using the alternative MF-CMA estimator for $H(q)$ to that obtained using the default MF-DFA estimator; results for the sample sizes of $N = 5,000$ and $N = 2,500$ are included, with the relevant values for the default MF-DFA estimator from Table 4.1 repeated here for ease of comparison. It can be seen that test statistic values obtained using the MF-CMA estimator for $H(q)$ and $\tau(q)$ also perform well, however the differences between the empirical size and the nominal size is typically larger for tests than for tests based on the default MF-DFA estimator, particularly for the sample size of 2,500. However, the empirical power when using the MF-CMA estimator is often higher than that obtained from the MF-DFA estimator, with these differences being even larger if differences in empirical size are taken into account.

Table 4.3: Empirical size and power comparison between MF-DFA & MF-CMA estimators

Nominal sig. level	Empirical size			Empirical power		
	0.10	0.05	0.01	0.10	0.05	0.01
N = 5,000						
MF-DFA estimator						
dH_{inf}	0.097	0.044	0.009	0.937	0.925	0.887
dH_{avg}	0.098	0.044	0.008	0.913	0.900	0.860
$d\tau_{\text{inf}}$	0.095	0.046	0.008	0.946	0.931	0.901
$d\tau_{\text{avg}}$	0.100	0.044	0.005	0.908	0.891	0.858
MF-CMA estimator						
dH_{inf}	0.097	0.049	0.012	0.954	0.946	0.921
dH_{avg}	0.089	0.052	0.012	0.931	0.918	0.891
$d\tau_{\text{inf}}$	0.099	0.049	0.012	0.961	0.952	0.930
$d\tau_{\text{avg}}$	0.091	0.054	0.011	0.927	0.914	0.889
N = 2,500						
MF-DFA estimator						
dH_{inf}	0.106	0.051	0.007	0.876	0.829	0.744
dH_{avg}	0.099	0.047	0.007	0.833	0.807	0.737
$d\tau_{\text{inf}}$	0.103	0.050	0.008	0.884	0.843	0.767
$d\tau_{\text{avg}}$	0.096	0.053	0.008	0.838	0.811	0.740
MF-CMA estimator						
dH_{inf}	0.086	0.043	0.009	0.880	0.850	0.792
dH_{avg}	0.090	0.043	0.011	0.855	0.819	0.756
$d\tau_{\text{inf}}$	0.085	0.041	0.009	0.897	0.871	0.806
$d\tau_{\text{avg}}$	0.090	0.047	0.009	0.850	0.827	0.762

4.5.4 Choice of Process Under the Null and Alternative

For the MRW process used to calculate empirical power under the alternative hypothesis, the strength of multifractality is controlled directly by the parameter λ^2 , with larger values of λ^2 corresponding to stronger multifractality and the limiting value of $\lambda^2 = 0$ corresponding to a unifractal process. Therefore, as $\lambda^2 \rightarrow 0$ the power of the tests should converge to the nominal significance levels and as $\lambda^2 \rightarrow \infty$ power should converge to unity. Table 4.4 and Figure 4.5 present the empirical power obtained for various values of λ^2 , including the default value of $\lambda^2 = 0.25$, to establish whether this convergence

holds in practice.

Table 4.4: Effect of the strength of multifractality on empirical power

Nominal sig. level	Empirical power		
	0.10	0.05	0.01
$\lambda^2 = 0.01$			
dH_{inf}	0.700	0.633	0.488
dH_{avg}	0.680	0.634	0.509
$d\tau_{\text{inf}}$	0.705	0.636	0.505
$d\tau_{\text{avg}}$	0.684	0.640	0.515
$\lambda^2 = 0.025$			
dH_{inf}	0.937	0.925	0.887
dH_{avg}	0.913	0.900	0.860
$d\tau_{\text{inf}}$	0.946	0.931	0.901
$d\tau_{\text{avg}}$	0.908	0.891	0.858
$\lambda^2 = 0.05$			
dH_{inf}	0.994	0.993	0.988
dH_{avg}	0.979	0.976	0.968
$d\tau_{\text{inf}}$	0.992	0.991	0.989
$d\tau_{\text{avg}}$	0.972	0.967	0.957

Table 4.4 reports empirical power values for all of the proposed test statistics with two additional values of $\lambda^2 = 0.05$ and $\lambda^2 = 0.01$ either side of the default value of 0.025 for a sample size of $N = 5,000$. Whilst the empirical power of the tests is already high at the default value of $\lambda^2 = 0.025$, increasing the strength of multifractality with a value of $\lambda^2 = 0.05$ results in even higher empirical power, with a minimum of nearly 96% for the $d\tau_{\text{avg}}$ statistic at a nominal significance level of 1%.

Whilst power does drop substantially when λ^2 is decreased from a value of 0.025 to 0.01, it can be seen from Figure 4.1(a) that the theoretical scaling function for the MRW process with a value of $\lambda^2 = 0.01$ is very close to the linear function of a unifractal process; over the range $0 \leq q \leq 5$ such weak multifractality would be nearly impossible to detect using the previous graphical approach and yet the proposed testing methodology still results in empirical power of approximately 70% for all of the test statistics at a

nominal significance level of 10% and approximately 50% at a nominal significance level of 1%.

Figure 4.5 plots empirical power as a function of multifractality strength¹⁶ for a wider range of values for λ^2 in order to give a more complete picture of the convergence of empirical power. Figure 4.5 contains plots just for the dH_{inf} and dH_{avg} statistics based on the generalised Hurst exponent, with the equivalent plots for the scaling function based statistics omitted since they were nearly identical to those presented. The convergence of empirical power to the limiting values in both directions is clearly visible from Figure 4.5, with a value of $\lambda^2 = 0.1$ producing a minimum empirical power 0.949 and thus almost complete convergence. Moving in the opposite direction, convergence to the nominal significance levels as $\lambda^2 \rightarrow 0$ is almost complete at the value of $\lambda^2 = 0.0001$, although convergence in this direction does appear to be somewhat slower.

To confirm that the tests have power against other multifractal process in addition to the MRW, the Markov-switching multifractal (MSM) model of Calvet and Fisher (2004) was also tested, which represents the most notable alternative to the MRW as a theoretical multifractal process for financial applications. The number of volatility components $\bar{k}$ for the MSM model was set to 5 and following the estimates obtained for financial data by Calvet and Fisher (2004), the parameter values used for simulation of the MSM process were $\sigma = 0.5, m_0 = 1.5, b = 8, \gamma_{\bar{k}} = 0.75$, which correspond to the standard deviation, binomial value, frequency growth rate and high-frequency switching probability respectively (see Calvet and Fisher, 2004 for further details).

Table 4.5 reports empirical power of the test statistics against the MSM alternative for sample sizes of 5,000 and 2,500, with all other aspects of the testing methodology chosen according to the default choices discussed in Section 4.5.1. It is clear from Table 4.5

¹⁶Note that for Figure 4.5 a logarithmic scale has been employed for the horizontal axis to improve legibility, however the vales of λ^2 marked on the axis correspond to the actual values of λ^2 and not their logarithms.

Table 4.5: Empirical power against Markov-switching multifractal alternative

Nominal sig. level	Empirical power		
	0.10	0.05	0.01
N = 5,000			
dH_{inf}	0.995	0.995	0.991
dH_{avg}	0.980	0.977	0.971
$d\tau_{\text{inf}}$	0.997	0.997	0.989
$d\tau_{\text{avg}}$	0.976	0.970	0.954
N = 2,500			
dH_{inf}	0.985	0.977	0.964
dH_{avg}	0.958	0.950	0.926
$d\tau_{\text{inf}}$	0.986	0.980	0.969
$d\tau_{\text{avg}}$	0.945	0.936	0.915

that the proposed tests maintain good power properties against the relevant MSM alternative with appropriate parameter values for financial data, actually resulting in higher empirical power values than against the default MRW alternative with default parameter values. However, given the large variation in empirical power observed previously for the default MRW process as multifractality strength is varied, without a universal measure to compare the strength of multifractality between each of these multifractal alternatives, direct comparisons of empirical power between the MSM and MRW processes are not possible.

Finally, the size and power properties of the test are assessed for series with long-range dependence in the level of the process. All series simulated thus far have been constructed to exhibit linear independence in the level of the series¹⁷, following the empirical finding that financial returns generally exhibit little to no serial correlation in their levels, despite the strong non-linear dependence typically observed.

The fGn process used to measure empirical size under the null exhibits linear dependence by construction for any value of the generalised Hurst exponent satisfying $H \neq 0.5$ and linear independence for $H = 0.5$. The MRW used to measure power under

¹⁷Although as previously discussed, the MRW process used for testing power under the alternative will exhibit non-linear serial dependence, particularly in the variance, even when linearly independent.

the alternative was generalised to allow for linear dependence by replacing the simple Gaussian noise $\epsilon_{\Delta t, k}$ of Section 4.5.1 with a fractional Gaussian noise with Hurst exponent $H \neq 0.5$, as described by Bacry et al. (2001). Given that the estimated linear dependence for return series is typically weak if found to be non-zero, attention is restricted to two additional values of $H = 0.6$ and $H = 0.4$, which correspond to weak linear persistence and weak linear anti-persistence respectively.

Table 4.6: Empirical size and power with linearly dependent series

Nominal sig. level	Empirical size			Empirical power		
	0.10	0.05	0.01	0.10	0.05	0.01
H = 0.4						
dH_{inf}	0.089	0.048	0.007	0.962	0.953	0.927
dH_{avg}	0.105	0.049	0.006	0.933	0.918	0.891
$d\tau_{\text{inf}}$	0.091	0.047	0.009	0.964	0.954	0.929
$d\tau_{\text{avg}}$	0.104	0.053	0.008	0.916	0.906	0.879
H = 0.5						
dH_{inf}	0.097	0.044	0.009	0.937	0.925	0.887
dH_{avg}	0.098	0.044	0.008	0.913	0.900	0.860
$d\tau_{\text{inf}}$	0.095	0.046	0.008	0.946	0.931	0.901
$d\tau_{\text{avg}}$	0.100	0.044	0.005	0.908	0.891	0.858
H = 0.6						
dH_{inf}	0.098	0.045	0.007	0.925	0.898	0.843
dH_{avg}	0.094	0.047	0.008	0.886	0.866	0.802
$d\tau_{\text{inf}}$	0.098	0.040	0.008	0.931	0.906	0.860
$d\tau_{\text{avg}}$	0.090	0.048	0.012	0.883	0.859	0.805

Table 4.6 reports empirical size and power for these additional values of H and it is clear from the reported results that there are no major changes in the empirical size and power of the tests compared to the case of linear independence, with all of the test statistics still performing well. In terms of empirical size, there appears to be a small drop in performance for the infimum-based statistics for the case of $H = 0.4$ and a drop for the average-based statistics for the case of $H = 0.6$. Empirical power appears to be mostly unaffected for the case of $H = 0.4$, but seems to fall slightly for all test statistics when $H = 0.6$, particularly for the average-based statistics. Nonetheless, given that

the estimated strength of linear dependence observed in asset return series is generally weaker than that implied by the values of $H = 0.6$ and $H = 0.4$, these effects are not likely to be problematic in practice when applying the testing methodology to real financial data.

4.6 Empirical Exercise

The current section contains a short empirical exercise applying the proposed testing methodology to intraday returns for several major financial assets in order to demonstrate the application of the tests to a real dataset. Given the dynamic estimation and forecasting framework of the previous chapters, particular attention is paid to tests of the local, rather than global, scaling properties of the return series.

The dataset used for the empirical exercise of the current chapter is identical to that employed in Chapters 2 and 3, both in terms of the raw dataset and the methods employed to process and prepare the data. Full details of the methods used for the preparation of the data can be found in Chapter 2 Section 2.5.1, with the key aspects summarised here for convenience.

The data contain intraday 5-minute observations from 3rd January 2007 until 31st December 2010 on the Euro (EUR) and Japanese Yen (JPY) exchange rates against the US Dollar (USD) and the levels of the S&P500 and NASDAQ-100 equity indexes. From these raw price series weekends and other non-trading days were removed; for the S&P500 and NASDAQ-100 data this consisted of removing all weekends and the non-weekend closures in historical list of holidays available on the NYSE website. For the exchange rate series, which have 24-hour trading 7 days a week, the 48-hour weekend periods between 21:05GMT on each Friday and 21:00 on each Sunday were removed, together with Christmas Day and New Year's Day, which were the only holidays during which trading was noticeably slower than normal. This process leaves a sample size

of 1008 trading days for the equity index series and 1037 for the exchange rate series. Continuous 5-minute returns were then constructed from the first difference of the log-price series for each asset, with the first 5-minute return for each day calculated between the closing price in the previous trading day and the opening price in the current day.

As discussed previously in Section 4.2.3, the scaling properties of all asset return series in the current dataset appear to be consistent with a multifractal process when assessed using the informal graphical testing method used previously in the finance literature; from Figure 4.3, all sample estimates of the scaling function $\tau(q)$ are strictly concave and equivalently all estimates of the generalised Hurst exponent are decreasing functions of q , both of which are consistent with a multifractal rather than unifractal process.

When applying the proposed testing methodology to the complete sample of 5-minute returns for each series, the null of unifractality is firmly rejected with p-values equal to zero to four decimal places, regardless of which of the proposed test statistics is selected. Whilst these results provide very strong evidence that the global scaling properties of these return series over the complete sample period are consistent with a multifractal rather than unifractal process, such a finding is expected based on the results of the informal graphical testing approach in Section 4.2.3; whilst this informal testing approach is potentially unreliable in smaller samples, it should perform acceptably in this situation given the very large sample sizes provided by the use of 5-minute intraday data, with over 250,000 and 75,000 observations for the exchange rate and equity index series respectively.

It is perhaps more interesting therefore to apply the proposed testing methodology to the problem of testing the local scaling properties of the return series over shorter periods of time, instead of the global scaling properties over the complete sample. The local scaling properties of the return series are also more relevant for the preceding work of Chapters 2 and 3, given that estimation of the scaling properties was performed

there in a dynamic context using rolling windows of data, rather than for the sample as a whole. In addition, as briefly mentioned in Section 4.2.3, it was previously demonstrated by Schumann and Kantelhardt (2011) that spurious multifractality may also be caused by so-called multifractional processes that are locally unifractal within distinct sub-periods, but with simple Hurst exponents that differ from one sub-period to another. This time-variation in the unifractal scaling properties of the series may make the scaling properties of the process appear multifractal when viewed globally, however this does not constitute true multifractal behaviour (since the process should be locally and globally multifractal) and so can be classified as a form of spurious multifractality.

The local scaling properties of the return series can be studied by splitting the sample of data into consecutive non-overlapping windows¹⁸ and then applying the proposed tests for unifractality versus multifractality to each of these sub-samples. Figure 4.6 plots the sample p-values obtained from applying the proposed test statistics to the first 250,000 5-minute returns for the EUR/USD data split into non-overlapping windows or sub-periods of lengths 2,500 (giving 100 windows) and 1,000 (giving 250 windows). The same exercise was also performed for the second JPY/USD exchange rate series, but the results obtained were very similar to those for the EUR/USD data and so equivalent figures have been omitted to conserve space.

For non-overlapping windows of length 5,000 or longer, as previously found for the complete sample of data the sample p-values for the null of unifractality within each sub-period remain equal to zero, again suggesting multifractal scaling. However, from Figure 4.6 it can be seen that once the window size is reduced to 2,500, for some sub-periods of the sample the null of unifractality can no longer be rejected. Decreasing the window size again to 1,000 further increases the number of sub-periods in which the null of unifractality cannot be rejected, with near identical results also observed for the second JPY/USD exchange rate series.

¹⁸This method could of course be modified to use a rolling window scheme with overlapping instead of non-overlapping windows.

The sub-period length for testing is not reduced below 1,000, because the theoretical properties of the testing methodology were not studied for such small sample sizes in the preceding Monte Carlo exercises, however it seems probable that the number of sub-periods exhibiting apparently unifractal properties will continue to increase as the length of each sub-period is reduced. One possible explanation for this finding is that the return series are multifractional rather than multifractal, exhibiting unifractal behaviour within shorter sub-periods of the complete sample, but with different simple Hurst exponents in different sub-periods, leading to spurious multifractality as discussed above. A second possibility is that the return process is truly multifractal in nature, both locally and globally, with the inability to reject the null of unifractality for some shorter sub-periods being due to the power of the tests decreasing as the sample size is reduced.

Figure 4.7 contains equivalent plots to those in Figure 4.6 for the S&P500 equity index data. The same consecutive non-overlapping window sizes of 2,500 and 1,000 are employed, however the smaller number of 5-minute observations for the equity index series mean that a smaller number of sub-periods are obtained for each window length. Additionally, as with the two exchange rate series, the two equity index series produced very similar results and so equivalent plots for the second equity index, the NASDAQ-100, have been omitted.

It is immediately obvious from Figure 4.7 that the results for the equity index data are very different to those previously reported for the exchange rate data, with weak evidence of local unifractality only beginning to appear once the sub-period length is reduced to 1,000. Given that the sub-period lengths used here are the same as those employed above for the exchange rate data, where the evidence of locally unifractality was much stronger, this casts some doubt on the previous explanation of limited power of the tests in finite samples. If all of the return processes were locally and globally multifractal, with the increase in apparent local unifractality for shorter sub-periods attributable to the power of the tests decreasing with the sample size, then it seems that

the observed local unifractality for the different series should be similar when using the same sub-period lengths. The fact that very different patterns are observed in practice between the exchange rate and equity index series suggests that other characteristics of the return processes are relevant, beyond the size of the sample that the tests are applied to.

In particular, although these sub-periods contain the same number of intraday observations, because of the large differences in trading hours between the foreign exchange and equity markets they correspond to very different lengths of trading time: the windows of 2,500 5-minute returns represent approximately 8.5 trading days for the exchange rate series, but approximately 30 trading days for the equity index series. Likewise the shorter windows of 1,000 5-minute returns correspond to approximately 3.5 trading days for the exchange rate and equity index series respectively. It is possible therefore that the degree of local unifractality observed is related not to the sub-period length in terms of the number of observations, but to the length of each sub-period in terms of trading time. This dependence on trading time rather than sample size would be more consistent with the explanation of a multifractional return process, with the lengths of each individually unifractal sub-period related in some way to trading time.

4.7 Conclusion

In response to the lack of a formal and generally applicable method for distinguishing between unifractal and multifractal scaling, the current chapter has presented a complete statistical testing methodology for determining whether a given sample of data is most consistent with a unifractal or multifractal process. The proposed testing methodology consists of a set of four possible statistics for testing the null hypothesis of unifractality against the alternative of multifractality, together with a model-based bootstrap resampling scheme that is used to obtain sample p-values and critical values for the statistics.

As with the informal testing methods employed previous in the literature, the set of test statistics are based directly on the differences in either the generalised Hurst exponent or the scaling function that exist between unifractal and multifractal processes. Crucially however, unlike these previous informal methods, the proposed model-based bootstrap resampling algorithm takes into account the estimation error for these functions, producing p-values or critical values that reflect the complete estimation method employed. Furthermore, the size of the deviation from unifractality measured by each of the statistics is calculated nonparametrically through the use of numerical differentiation methods, thus avoiding the need to specify a particular multifractal process under the alternative hypothesis.

In a series of Monte Carlo exercises using simulated unifractal and multifractal data, the proposed testing methodology is demonstrated to have good size and power properties in a wide range of situations and when using any of the four proposed test statistics. Strong performance is maintained even in sample sizes that would be considered to be very small in the literature on unifractal and multifractal processes, thus allowing tests of the local scaling properties of the process of interest within short sub-periods of the complete sample of data.

In addition, the methodology is shown to be robust to changes in the estimation method employed to obtain the values of the generalised Hurst exponent or scaling function from which the test statistics are calculated. This suggests that, as intended, the resampling scheme developed does successfully take into account the effects of the chosen estimation method when producing p-values and critical values for the tests.

Varying the strength of multifractality for the multifractal random walk process used to estimate empirical power under the alternative demonstrates that the high power of the tests is maintained even for multifractality strengths that are lower than the most plausible range of values for financial data. The power of the tests also remains high when the original multifractal random walk under the alternative is exchanged for the Markov-

switching multifractal model, thus demonstrating consistent performance against the two most plausible theoretical multifractal processes proposed in the literature for financial applications.

The testing methodology is also applied to the same dataset of intraday exchange rate and equity index returns used for the empirical exercises of the earlier chapters. Applying the test to the complete sample of intraday returns for each asset the null of unifractality is strongly rejected in all cases, implying that the data are more consistent with multifractal distributional scaling, at least in a global sense. Given the strong focus on local scaling properties in earlier chapters and the seemingly good performance of the testing methodology for smaller samples, the tests are also applied to shorter sub-periods of the complete intraday return series.

Evidence of local unifractality is found for some sub-periods, with the strength of this apparent local unifractality increasing as the length of these sub-periods is decreased. However, the large differences between the results for the exchange rate and equity index data for sub-periods containing the same number of observations suggests that this apparent local unifractality is related not to the sample size of the sub-periods, but their length in terms of trading time. It seems therefore that this apparent local unifractality cannot be attributable solely to the power of the tests decreasing in smaller samples, but is instead related to the distributional scaling properties of the different return series.

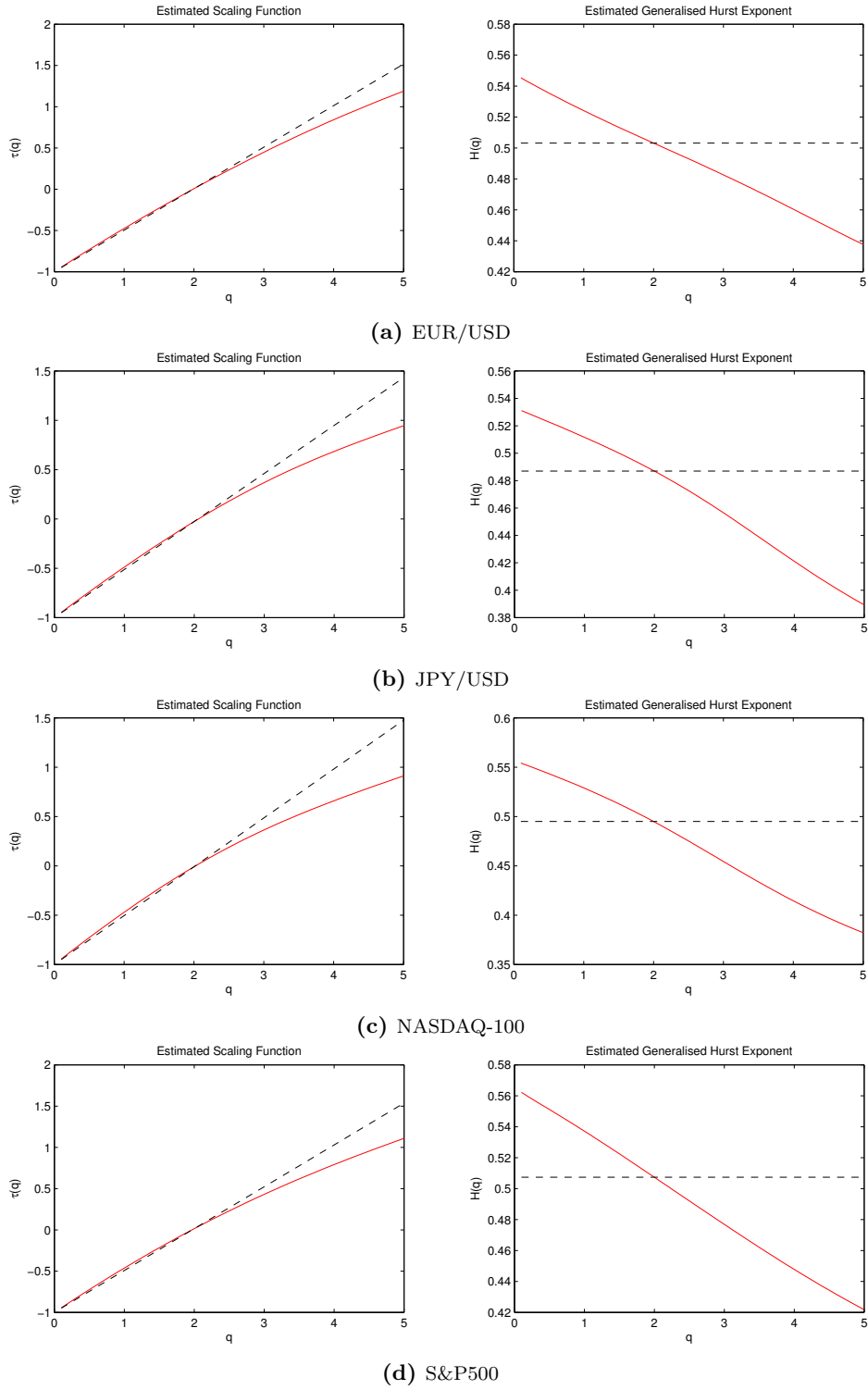
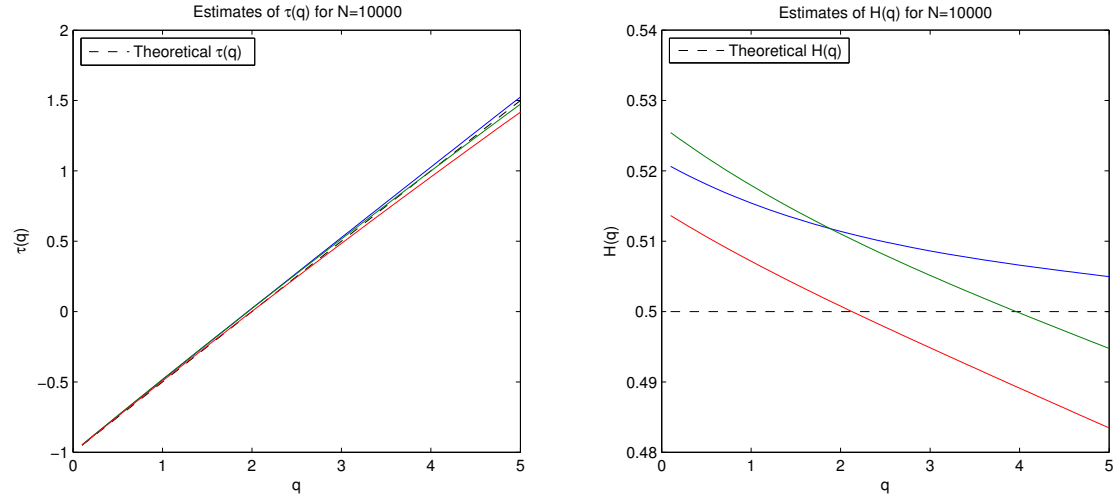
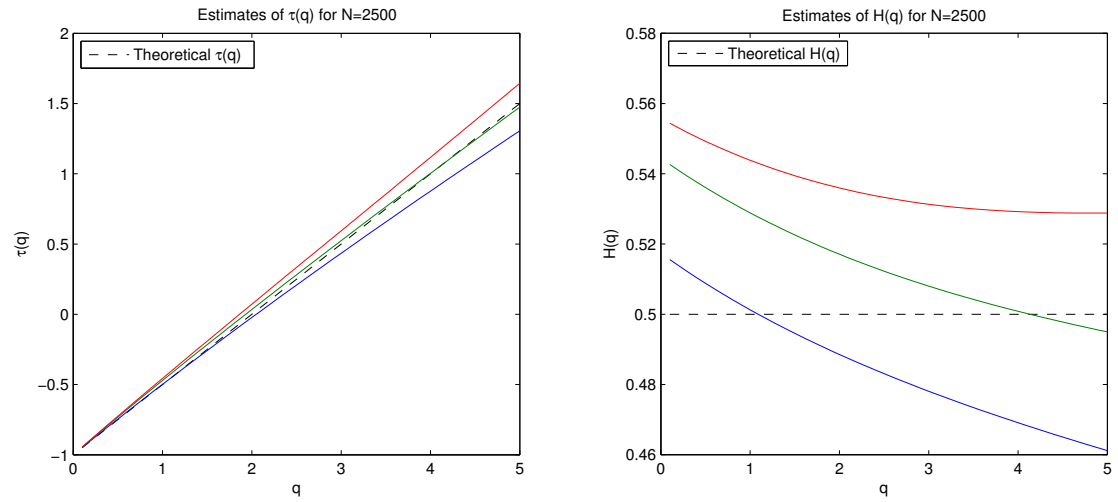


Figure 4.3: Estimated values of the scaling function and generalised Hurst exponent for financial data. Estimates are obtained using the MF-DFA estimator with linear detrending.

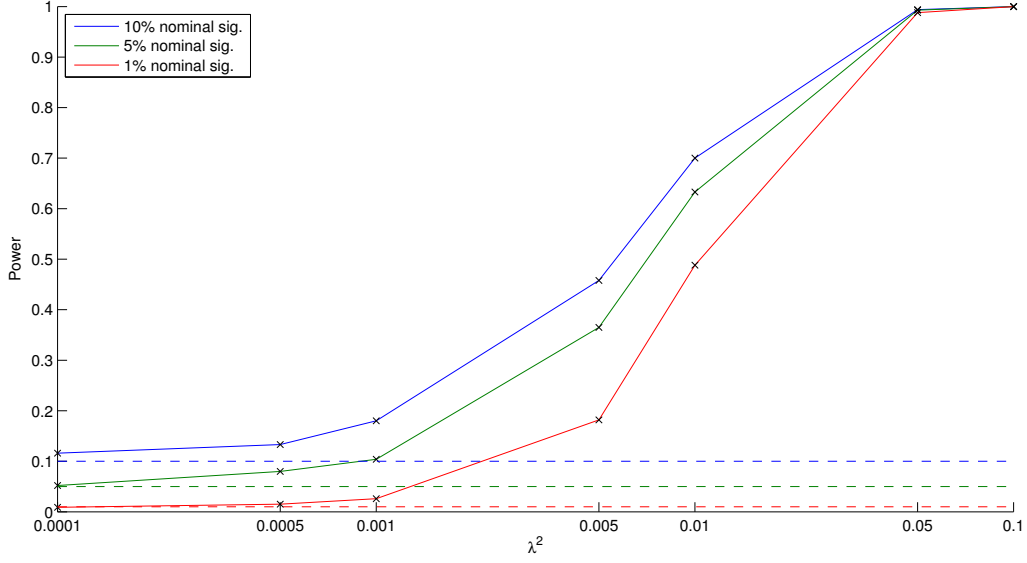


(a) Estimated scaling functions and generalised hurst exponents - $N = 10,000$

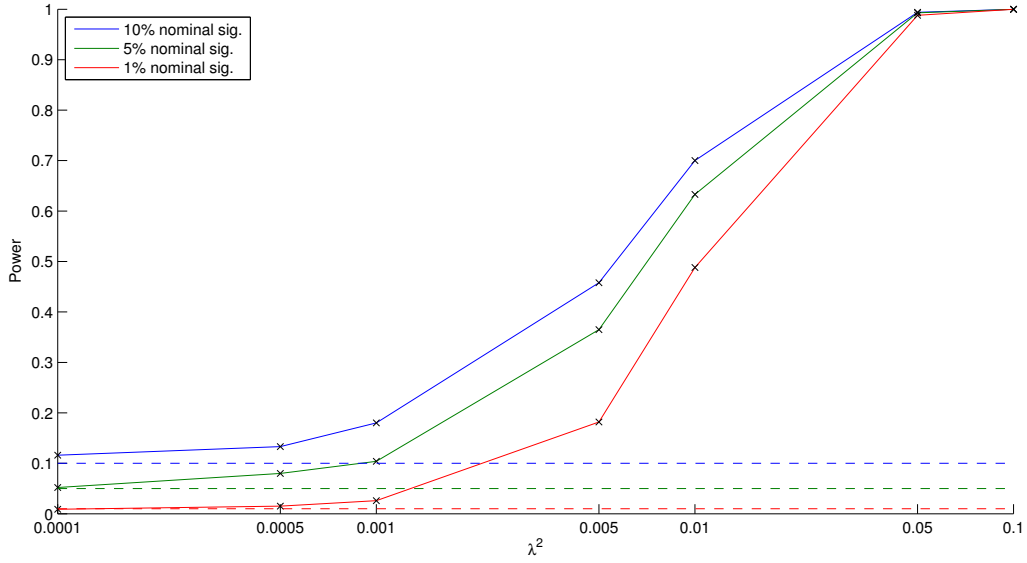


(b) Estimated scaling functions and generalised hurst exponents - $N = 2,500$

Figure 4.4: Estimates of the scaling function and generalised Hurst exponent for unifractal fractional Gaussian noise realisations. Dashed lines show the theoretical values for each of the functions.

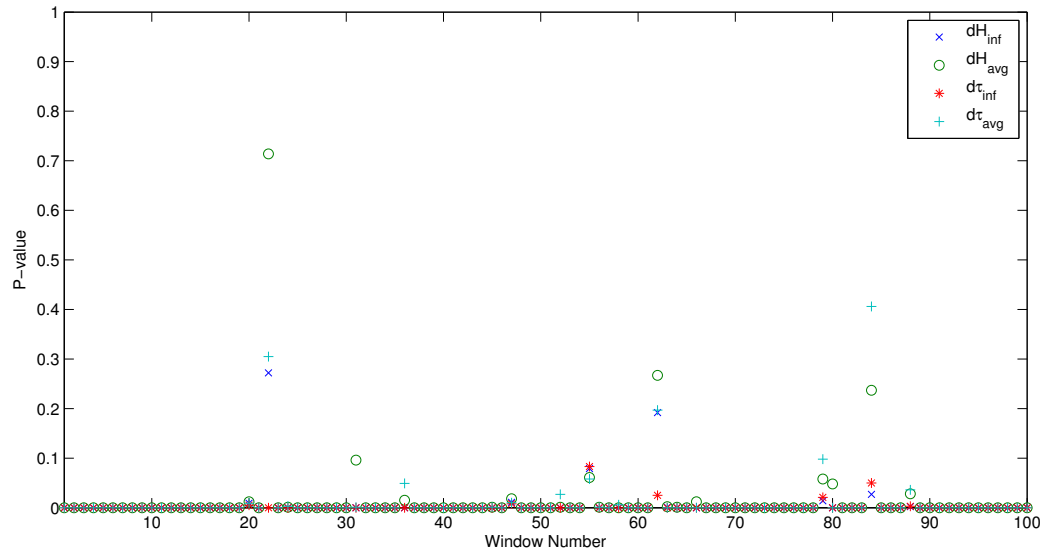


(a) Empirical power for dH_{inf} statistic

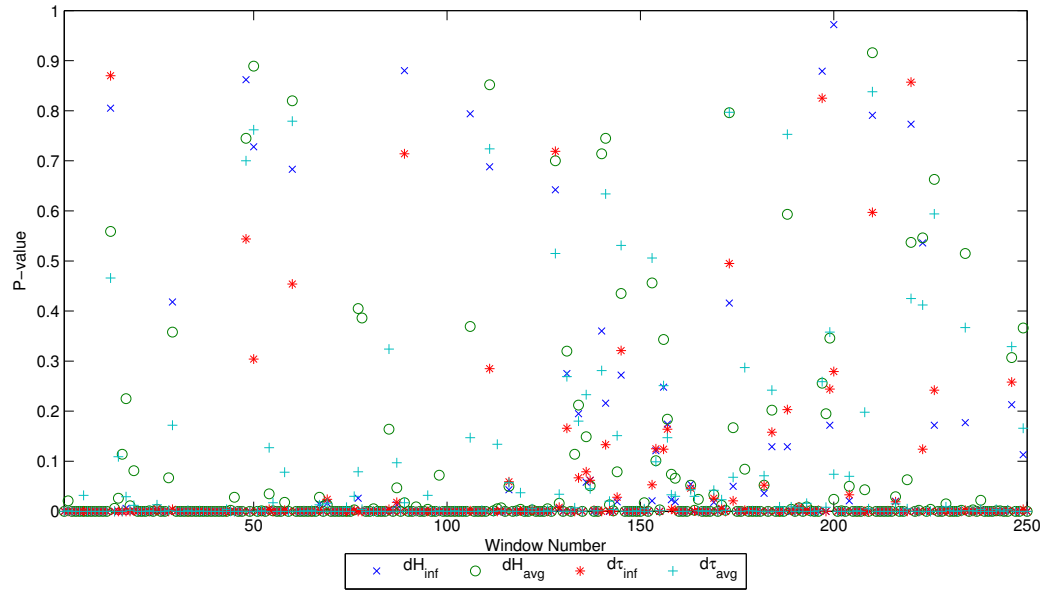


(b) Empirical power for dH_{avg} statistic

Figure 4.5: Empirical power versus multifractality strength (λ^2) for dH_{inf} and dH_{avg} statistics. A logarithmic scale has been used for the horizontal axis to improve legibility, but axis values correspond to true values of λ^2 and not logarithmic values.

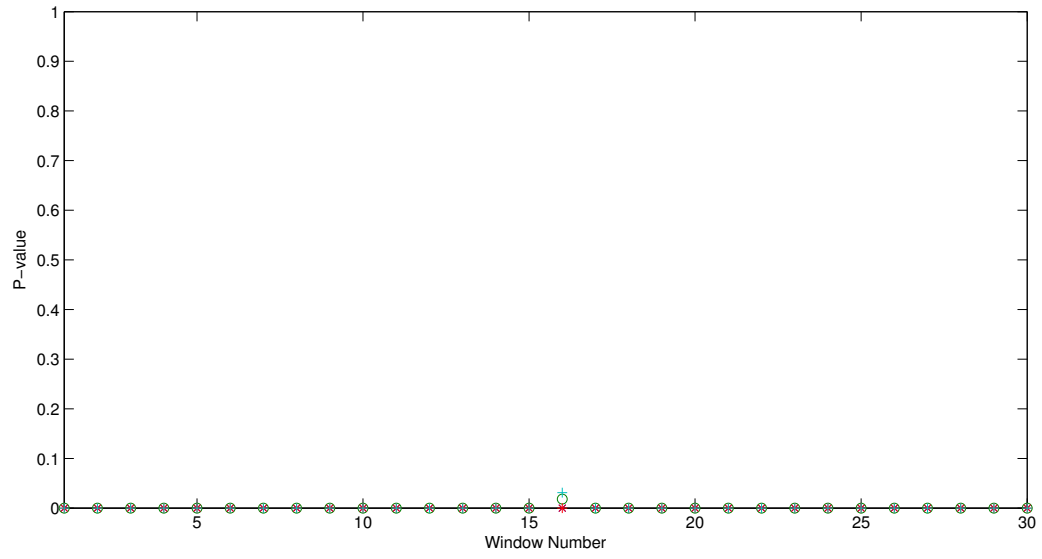


(a) Consecutive non-overlapping sub-periods of length 2,500

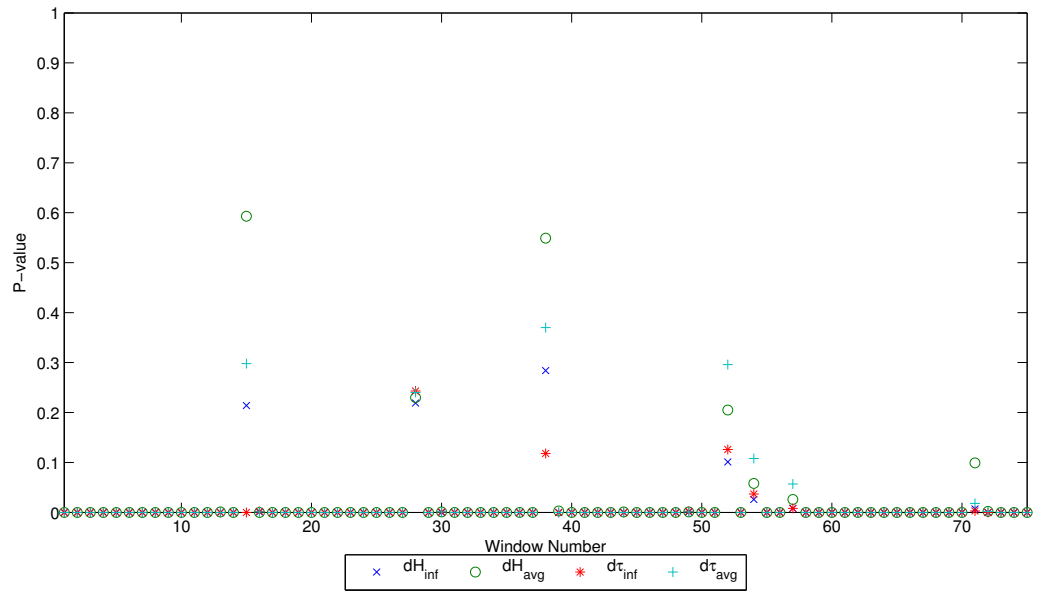


(b) Consecutive non-overlapping sub-periods of length 1,000

Figure 4.6: P-values for tests of unifractality versus multifractality applied to consecutive non-overlapping sub-periods of 5-minute EUR/USD data



(a) Consecutive Non-overlapping Sub-periods of Length 2,500



(b) Consecutive non-overlapping sub-periods of length 1,000

Figure 4.7: P-values for tests of unifractality versus multifractality applied to consecutive non-overlapping sub-periods of 5-minute S&P500 data

Chapter 5

Concluding Remarks

This thesis has attempted to address some of the current gaps in the growing literature on unifractal and multifractal processes in finance, through a combination of empirical and theoretical contributions spanning the key problems of estimation, forecasting and inference.

Chapters 2 and 3 proposed new methods for producing density forecasts for daily returns from intraday data under the assumption of unifractality and multifractality respectively. This is achieved through new applications of the distributional scaling laws satisfied by unifractal and multifractal processes, which provide a new method of incorporating intraday information into the estimation and forecasting of daily return densities. At the same time, the use of a dynamic estimation environment in Chapters 2 and 3 represents a secondary contribution to the literature. By allowing the distributional scaling properties to be time-varying, additional flexibility is introduced compared to the vast majority of work in the existing literature, in which a static estimation environment is imposed with the scaling properties of the return process assumed to be fixed for the complete sample period.

In Chapter 2, it was demonstrated how a new application of the theoretical distributional scaling properties for the simpler class of unifractal processes can be employed

in practice to estimate the density function of daily returns directly from appropriately rescaled intraday observations. Whilst the structure of distributional scaling implied by the assumption of unifractality is more restrictive than that in the multifractal case, in practice it is sufficient for unifractal scaling to be present locally within each trading day and only over the range of sampling intervals of direct interest, rather than over all timescales and across any sub-period of the data as for a true unifractal process.

One of the key theoretical advantages of the proposed unifractal method over existing methods to estimate and forecast daily return densities from intraday data is that it permits the use of nonparametric specifications for the distribution of daily returns and preserves information about both the sign and magnitude of the intraday returns. However, the use of nonparametric specifications for the daily return density complicates the task of forecasting, by rendering conventional approaches based on imposing dynamics via the distributional parameters inapplicable. Instead, a new approach is proposed inspired by results and methods from the literature on density forecast combination, that imposes a parametric dynamic structure directly onto the estimated return densities.

The out-of-sample density forecasting performance of the unifractal method was compared against existing methods using both intraday and daily data in an empirical application with 5-minute intraday data on major exchange rates and equity indexes. The predictive ability of the unifractal approach is shown to be competitive with existing methods based on intraday and daily data, particularly for shorter in-sample periods and during periods of high return volatility. However, as expected the relative performance of the method is stronger for return series with distributional scaling properties close to the unifractal scaling required. Forecasting performance is poorer, though still competitive, for time series that exhibit larger deviations from unifractality and possess scaling properties seemingly closer to that of the more general multifractal case.

Given the apparent limitations of the unifractal forecasting approach of Chapter 2 when applied to return series that deviate more substantially from the required unifractal

tal scaling, an equivalent density forecasting method was developed in Chapter 3 under the more general assumption of multifractal distributional scaling. Whilst this assumption permits greater flexibility in the structure of distributional scaling than for the unifractal processes of Chapter 2, the more general distributional scaling laws are less straightforward to apply in practice, resulting in a loss of flexibility elsewhere. Most notably nonparametric specifications can no longer be used for the daily return density and only the moments of the daily returns can be estimated from the intraday data.

The predictive ability of the multifractal density forecasting approach is evaluated in an equivalent empirical exercise to that of Chapter 2, with the performance of the method compared to both existing methods from the literature and the unifractal approach of the preceding chapter. For the EUR/USD data the multifractal method provides large improvements in predictive ability over the GARCH benchmark model and is competitive with realised volatility based methods. For the remaining return series, the null of equal predictive ability is unable to be rejected in the majority of cases, implying that the density forecasting performance of the multifractal approach is competitive with the existing GARCH and realised volatility methods from the literature. A potential explanation for this variation in forecasting performance across the assets is provided by differences in the strength of distributional scaling for the return series, with the performance of the multifractal method better for return series exhibiting seemingly stronger distributional scaling.

Compared to the unifractal density forecasting method of Chapter 2, the predictive ability of the multifractal method is typically weaker, particularly for the equity index data where the unifractal method consistently provides highly significant gains in predictive ability over the multifractal method across the whole domain of the return density. This suggests that in the case of equity index data, the ability to employ nonparametric specifications for the daily return density is more beneficial than the additional flexibility in distributional scaling properties permitted by moving from a unifractal to a multi-

fractal context. The same is not true however for the exchange rate data, where the relative performance of the unifractal and multifractal methods is substantially closer, with the null of equal predictive ability not able to be rejected in most cases.

Finally, the purely statistical density forecast evaluation criteria employed so far are augmented by applying the various forecasting methods to the problem of portfolio allocation, in which the competing density forecasts are employed to optimally allocate funds between a risky and risk-free asset. In this context it is again found that the new multifractal approach can provide substantial gains over existing methods for the EUR/USD data, when measured in terms of the certainty equivalent return of the resulting portfolio. For the other assets the multifractal approach is found to outperform the realised volatility based method for higher levels of investor risk aversion.

In Chapter 4, a formal testing framework is proposed for determining whether a given sample of data is most consistent with a unifractal or multifractal data generating process, motivated by the importance of this issue and the lack of an effective test for this purpose in the existing literature.

The first key contribution of the proposed testing methodology is to develop a set of statistics for testing the null hypothesis of unifractality against the alternative of multifractality. As with existing informal graphical testing methods employed in the literature, the theoretical basis for these test statistics is provided by the differences that exist in the functions characterising the distributional scaling properties for unifractal and multifractal processes. Crucially however, these statistics provide a formal measure of the strength of multifractality exhibited by the sample estimates of these functions and furthermore, do not require any assumptions to be imposed on the form of multifractality under the alternative hypothesis due to the use of numerical differentiation methods.

Unfortunately, the distributions of the proposed test statistics are non-standard and the relevant rates of convergence are unknown, due to the specific characteristics of the testing environment and the complex theoretical properties of unifractal and multifractal

processes. The second contribution of the testing framework is therefore to develop an appropriate model-based bootstrap resampling scheme for approximating the distributions of the proposed test statistics under the null of unifractality in order to calculate critical values or p-values for the tests.

The size and power properties of the proposed testing procedure are evaluated in a series of Monte Carlo exercises using simulated unifractal and multifractal data, representing a wide range of situations in which the methodology may be applied in practice. These exercises demonstrate that the proposed testing methodology possesses good size and power properties in wide range of situations and appears to be robust to changes in either the implementation of the tests or the properties of the true data generating process under the null and alternative hypotheses. In particular, the tests perform well for sample sizes that would be considered as small in the multifractality literature, thus confirming the suitability of the methodology for the study of both local and global scaling properties, or in situations where data availability is limited.

Finally, the testing methodology is applied to the same dataset of intraday exchange rate and equity index data used in the previous chapters in order to examine both the global and local scaling properties of the data. Whilst the global scaling properties of all the return series are found to be consistent with multifractal rather than unifractal scaling behaviour, the test is also applied to shorter sub-periods of the complete intraday return series in order to study the scaling properties on a more local scale, which is arguably more relevant for the work of the previous chapters. Evidence of local unifractality is found for some sub-periods, with the strength of this apparent local unifractality increasing as the length of these sub-periods is decreased. Interestingly however, this apparent local unifractality does not seem to be explainable solely by a lack of power of the tests in smaller samples, but appears instead to be related to the fundamental distributional scaling properties of the different return series when examined over shorter sub-periods.

The literature on unifractal and multifractal processes in finance is still relatively young and evolving rapidly, with numerous possible directions for future research following both from the work of this thesis and also from elsewhere in the literature.

The first is the increasing interest in multivariate scaling, which allows for the possibility that the cross-correlations between different time series also exhibit scaling behaviour across sampling intervals. Work in this area has included efforts both to extend existing univariate theoretical models to the multivariate case, such as the bivariate extension of the Markov-switching multifractal model proposed by Calvet et al. (2006), and also to develop new methods to estimate the structure of this multivariate scaling in an empirical context (see for example Jiang and Zhou, 2011).

Most obviously, the forecasting methods of Chapters 2 and 3 could potentially be extended to a multivariate context using the methods under development in the literature, allowing more general financial applications to be considered by jointly forecasting the return series for multiple assets. The testing framework of Chapter 4 could however also potentially be extended to a multivariate setting, in order to test for the presence of distributional scaling between series.

Another interesting but currently undeveloped branch of the literature concerns the application of these scaling properties to the prediction of financial crises and crashes. This is based on the empirical observation that the structure of distributional scaling appears to change substantially in periods preceding financial crises (see for example Grech and Mazur, 2004, Stavroyiannis et al., 2011, or Siokis, 2013).

Whilst the initial empirical findings in this area are promising, the analysis performed so far is somewhat limited; in particular, previous studies do not adequately explore how best to incorporate these measures into crisis prediction methods and do not formally investigate the predictive ability of these new measures in isolation, or relative to existing leading indicators from the literature. As such, further work is necessary to establish the potential value of these measures of scaling behaviour as leading indicators for predicting

financial crises. In this context the testing framework of Chapter 4 would be particularly applicable, given the need to study the local scaling properties of the series in a dynamic context.

Bibliography

- E. Alessio, A. Carbone, G. Castelli, and V. Frappietro. Second-order moving average and scaling of stochastic time series. *The European Physical Journal B-Condensed Matter and Complex Systems*, 27(2):197–200, 2002.
- J. Alvarez-Ramirez, E. Rodriguez, and J. Carlos Echeverría. Detrending fluctuation analysis based on moving average filtering. *Physica A*, 354(0):199–219, 2005.
- G. Amisano and R. Giacomini. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, 25(2):177–190, 2007.
- T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55, 2001.
- T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.
- E. Bacry, J. Delour, and J. F. Muzy. Multifractal random walk. *Physical Review E*, 64(2):026103, 2001.
- E. Bacry, A. Kozhemyak, and J. F. Muzy. Continuous cascade models for asset returns. *Journal of Economic Dynamics and Control*, 32(1):156–199, 2008.

- T. Bao, T. Lee, and B. Saltoğlu. Comparing density forecast models. *Journal of Forecasting*, 26(3):203–225, 2007.
- A. Bashan, R. Bartsch, J. W. Kantelhardt, and S. Havlin. Comparison of detrending methods for fluctuation analysis. *Physica A: Statistical Mechanics and its Applications*, 387(21):5080–5090, 2008.
- C. Bender, T. Sottinen, and E. Valkeila. Arbitrage with fractional Brownian motion? *Theory of Stochastic Processes*, 13(29):23–34, 2007.
- S. Bianchi, A. Pantanella, and A. Pianese. Modeling stock prices by multifractional Brownian motion: an improved estimation of the pointwise regularity. *Quantitative Finance*, 13(8):1317–1330, 2013.
- J. Bouchaud, M. Potters, and M. Meyer. Apparent multifractality in financial time series. *The European Physical Journal B-Condensed Matter and Complex Systems*, 13(3):595–599, 2000.
- L. Calvet and A. Fisher. Multifractality in asset returns: theory and evidence. *Review of Economics and Statistics*, 84(3):381–406, 2002.
- L. Calvet and A. Fisher. How to forecast long-run volatility: regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics*, 2(1):49–83, 2004.
- L. Calvet, A. Fisher, and S. Thompson. Volatility comovement: a multifrequency approach. *Journal of Econometrics*, 131(1-2):179–215, 2006.
- F. Chen, F. X. Diebold, and F. Schorfheide. A Markov-switching multifractal inter-trade duration model, with application to US equities. *Journal of Econometrics*, 2013. doi: 10.1016/j.jeconom.2013.04.016.

- M. P. Clements, A. Galvão, and J. Kim. Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance*, 15(4):729–750, 2008.
- D. Delignieres, S. Ramdani, L. Lemoine, K. Torre, M. Fortes, and G. Ninot. Fractal analyses for ‘short’ time series: a re-assessment of classical methods. *Journal of Mathematical Psychology*, 50(6):525–544, 2006.
- T. Di Matteo. Multi-scaling in finance. *Quantitative finance*, 7(1):21–36, 2007.
- T. Di Matteo, T. Aste, and M. Dacorogna. Long-term memories of developed and emerging markets: using the scaling analysis to characterize their stage of development. *Journal of Banking & Finance*, 29(4):827–851, 2005.
- C. R. Dietrich and G. N. Newsam. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, 18(4):1088–1107, 1997.
- R. Dittmar. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *Journal of Finance*, 57(1):369–403, 2002.
- J. Duchon, R. Robert, and V. Vargas. Forecasting volatility with the multifractal random walk model. *Mathematical Finance*, pages no–no, 2010.
- R. Engle and S. Manganelli. CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381, 2004.
- J. Fillol. Multifractality: theory and evidence, an application to the French stock market. *Economics Bulletin*, 3(31):1–12, 2003.
- M. Frezza. Modeling the time-changing dependence in stock markets. *Chaos, Solitons and Fractals*, 45(12):1510–1520, 2012.
- R. Giacomini and H. White. Tests of conditional predictive ability. *Econometrica: Journal of the Econometric Society*, 74(6):1545–1578, 2006.

- P. Giot and S. Laurent. Modelling daily Value-at-Risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, 11:379–398, 2004.
- T. Gneiting and R. Ranjan. Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–244, 2011.
- T. Gneiting and M. Schlather. Stochastic models that separate fractal dimension and the hurst effect. *SIAM Review*, 46(2):269–282, 2004.
- D. Grech and Z. Mazur. Can one make any crash prediction in finance using the local Hurst exponent idea? *Physica A*, 336(1):133–145, 2004.
- D. Grech and G. Pamuła. Multifractal background noise of monofractal signals. *Acta Phys. Pol. A*, 121:B34–B39, 2012.
- S. Hall and J. Mitchell. Combining density forecasts. *International journal of forecasting*, 23(1):1–13, 2007.
- P. R. Hansen and A. Lunde. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, 24(2):127–161, 2006.
- C. Harvey and A. Siddique. Conditional skewness in asset pricing tests. *Journal of Finance*, 55(3):1263–1295, 2000.
- S. Jaffard, B. Lashermes, and P. Abry. Wavelet leaders in multifractal analysis. *Wavelet analysis and applications*, pages 201–246, 2007.
- Z. Q. Jiang and W.-X. Zhou. Multifractal detrending moving-average cross-correlation analysis. *Physical Review E*, 84(1):016106, 2011.
- E. Jondeau and M. Rockinger. Gram-Charlier densities. *Journal of Economic Dynamics and Control*, 25(10):1457–1483, 2001.

- J. Kantelhardt. Fractal and multifractal time series. In *Encyclopedia of Complexity and Systems Science*. Springer, 2009.
- J. Kantelhardt, S. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(1-4):87–114, 2002.
- C. Kascha and F. Ravazzolo. Combining inflation density forecasts. *Journal of Forecasting*, 29(1-2):231–250, 2010.
- B. Lashermes, P. Abry, and P. Chainais. New insights into the estimation of scaling exponents. *International Journal of Wavelets, Multiresolution and Information Processing*, 2:497–524, 2005.
- Q. Li and J. Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 2006.
- J. Ludescher, M. I. Bogachev, J. W. Kantelhardt, A. Y. Schumann, and A. Bunde. On spurious and corrupted multifractality: the effects of additive noise, short-term memory and periodic trends. *Physica A: Statistical Mechanics and its Applications*, 390(13):2480–2490, 2011.
- T. Lux. Detecting multifractal properties in asset returns: the failure of the “scaling estimator”. *International Journal of Modern Physics C*, 15(04):481–491, 2004.
- J. Maheu and T. McCurdy. Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics*, 160(1):69–76, 2010.
- B. Mandelbrot, A. Fisher, and L. Calvet. A multifractal model of asset returns. *Cowles Foundation Discussion Papers*, no. 1164, Available at SSRN: <http://ssrn.com/abstract=78588>, 1997.

- J. Mielniczuk and P. Wojdylo. Estimation of Hurst exponent revisited. *Computational Statistics & Data Analysis*, 51(9):4510–4525, 2007.
- J. Mitchell and S. Hall. Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR fan charts of inflation. *Oxford Bulletin of Economics & Statistics*, 67(1):995–1033, 2005.
- J. F. Muzy, E. Bacry, and A. Arneodo. Wavelets and multifractal formalism for singular signals: application to turbulence data. *Physical review letters*, 67(25):3515–3518, 1991.
- J. F. Muzy, D. Sornette, J. Delour, and A. Arneodo. Multifractal returns and hierarchical portfolio theory. *Quantitative Finance*, 1(1):131–148, 2001.
- E. Onali and J. Goddard. Unifractality and multifractality in the Italian stock market. *International review of financial analysis*, 18(4):154–163, 2009.
- A. Pagan. Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, 25(1):221–247, 1984.
- E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.
- F. Schmitt, D. Schertzer, and S. Lovejoy. Multifractal analysis of foreign exchange data. *Applied Stochastic Models and Data Analysis*, 15(1):29–53, 1999.
- A. Schumann and J. Kantelhardt. Multifractal moving average analysis and test of multifractal model with tuned correlations. *Physica A*, 390(14):2637–2654, 2011.
- F. Selçuk and R. Gençay. Intraday dynamics of stock market returns and volatility. *Physica-Section A*, 367:375–387, 2006.
- F. Siokis. Multifractal analysis of stock exchange crashes. *Physica A*, 392(5):1164–1171, 2013.

- S. Stavroyiannis, V. Nikolaidis, and I. Makris. On the multifractal properties and the local multifractality sensitivity index of euro to Japanese yen foreign currency exchange rates. *Global Business and Economics Review*, 13(1):93–103, 2011.
- A. Timmermann. Forecast combinations. In *Handbook of Economic Forecasting, Volume I*. Elsevier, 2006.
- F. Žikeš. Semiparametric conditional quantile models for financial returns and realized volatility. *Working Paper*, 2009.
- H. Wendt and P. Abry. Multifractality tests using bootstrapped wavelet leaders. *IEEE Transactions on Signal Processing*, 55(10):4811–4820, 2007.
- Z. Xu and R. Gençay. Scaling, self-similarity and multifractality in FX markets. *Physica-Section A*, 323:578–590, 2003.