

Statistical Tests of Distributional Scaling Properties for Financial Return Series

Mark Hallam;^{*†}
University of Essex

Jose Olmo
University of Southampton

*Corresponding author: Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, England. E-mail: mark.hallam@essex.ac.uk, Tel: +44 1206 873164.

[†]Mark Hallam would like to thank Science Foundation Ireland (under grant number 08/SRC/FMC1389) and The Scientific and Technological Research Council of Turkey (under the BİDEB 2216 Research Fellowship Programme for International Researchers, undertaken at Koç University) for financial support. The authors would also like to thank participants at the 7th CSDA International Conference on Computational and Financial Econometrics and anonymous referees for constructive feedback and comments.

Abstract

Existing empirical evidence of distributional scaling in financial returns has helped motivate the use of multifractal processes for modelling return processes. However, this evidence has relied on informal tests that may be unable to reliably distinguish multifractal processes from other related classes. The current paper develops a formal statistical testing procedure for determining which class of fractal process is most consistent with the distributional scaling properties in a given sample of data. Our testing methodology consists of a set of test statistics, together with a model-based bootstrap resampling scheme to obtain sample p-values. We demonstrate in Monte Carlo exercises that the proposed testing methodology performs well in a wide range of testing environments relevant for financial applications. Finally, the methodology is applied to study the scaling properties of a dataset of intraday equity index and exchange rate returns. The empirical results suggest that the scaling properties of these return series may be inconsistent with purely multifractal processes.

Keywords: Multifractal, Unifractal, Bootstrap inference, Hypothesis testing, Financial returns

1 Introduction

The class of multifractal processes have been applied to the problem of modelling complex systems in numerous fields, such as signal processing and the physical sciences, however in recent years these processes have attracted particular attention in the finance literature, where they have been proposed as an alternative class of time series process for modelling financial asset return series.

Several theoretical multifractal processes have been developed that are intended specifically for modelling asset returns, with key examples including the multifractal model of asset returns of Mandelbrot et al. (1997) and Calvet and Fisher (2002), the Markov-switching multifractal (MSM) model of Calvet and Fisher (2004), Calvet et al. (2006) and Lux and Morales-Arias (2010) and the multifractal random walk (MRW) of Bacry et al. (2001) and Muzy et al. (2001). These multifractal processes have been shown to provide flexible and parsimonious models for asset returns that are able to reproduce the central empirically observed properties of return series, such as volatility clustering. Furthermore, Chen et al. (2013) have also recently proposed a modified MSM model for inter-trade duration, suggesting that multifractal processes may have financial applications beyond the modelling of returns.

Multifractal processes may be defined in terms of distributional scaling properties, which implies that the statistical properties of the process at different sampling intervals or timescales are formally linked according to some theoretical scaling laws. In finance this distributional scaling behaviour appears to be particularly relevant for return series, with numerous empirical studies finding evidence that returns for a wide range of financial assets exhibit scaling properties consistent with these processes, including Schmitt et al. (1999), Calvet and Fisher (2002), Fillol (2003), Xu and Gençay (2003), Di Matteo et al. (2005), Di Matteo (2007) and Onali and Goddard (2009) and many others. This apparent empirical support for the presence of multifractal distributional scaling in return series has often been viewed as a confirmation of the appropriateness of multifractal processes for modelling return series and has helped motivate work on the theoretical side of the multifractal finance literature.

Unfortunately, the lack of a formal statistical test for assessing the presence or type of distributional scaling in a given sample of data means that the existing empirical evidence of multifractal distributional scaling in asset returns has been based instead on an informal graphical approach. It has been demonstrated that this graphical testing approach may not be able to reliably distinguish between the distributional scaling of a multifractal process and that of other closely related processes (see, for example, Bouchaud et al., 2000, Lux, 2004 and Grech and Pamuła, 2012). Most notably, simulated data from the simpler class of unifractal processes (which includes the standard and fractional Brownian motions, plus certain types of Levy process) may exhibit apparent multifractal distributional scaling when

tested using this approach, despite possessing a much more restrictive form of distributional scaling behaviour in reality.

Given the known limitations of the standard graphical test for distributional scaling, there is a clear need to develop more formal and reliable testing procedures for distinguishing between multifractal distributional scaling behaviour and that possessed by unifractal or other related classes of process. Little effort has however been made in this area, with the most notable exception in the context of finance being the work of Lux (2004) who proposes a formal statistical test for testing the null hypothesis of ‘no multifractality’ versus the alternative of multifractality. However, restrictions are implicitly imposed that limit the generality of the proposed testing approach; firstly, a specific multifractal process must be chosen under the alternative hypothesis and therefore the method cannot be viewed as a test for multifractality of a general or unspecified form, but is instead a test for distributional scaling consistent with a specific multifractal process. Secondly, the resampling method employed to obtain the distribution of the test statistics does not permit serial dependence of any form under the null hypothesis of no multifractality. Given that the majority of unifractal processes, or indeed any other process suitable for modelling return series, may exhibit some form of serial dependence, the null hypothesis seems inappropriate in applications to financial returns.

The only existing formal testing procedure that provides an explicit test for multifractality versus unifractality appears to be the wavelet-based approach of Wendt and Abry (2007), which was published in the signal-processing literature and has not been applied in the context of finance¹. However, in its proposed form the test imposes some additional constraints on the form of distributional scaling under the alternative of multifractality, thus somewhat limiting the generality of the test. More critically, it is clear both from the Monte Carlo exercises of Wendt and Abry (2007) and those performed in the current work that the wavelet-based tests require very 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 unavoidable, due either to limited data availability or the desire to study the local distributional scaling properties of a process over shorter sample periods.

In response to this gap in the literature, the current work 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

¹On the more financial and econometric side of the literature, wavelet-based methods have been widely applied to develop statistical tests in many other contexts, such for serial correlation (Gençay and Signori, 2015) and unit roots (Fan and Gençay, 2010), in addition to more empirical applications, such as hedging (Conlon et al., 2016).

under the alternative hypothesis. Furthermore, we focus on developing a testing procedure that exhibits good performance in sample sizes that although large by conventional econometric standards, are small by the standards of the literature on fractal processes. This permits the use of our tests for situations in which maximum sample size is limited or for tests of local distributional scaling properties as mentioned above.

We propose a set of possible test statistics for testing the null hypothesis of unifractality against the alternative of multifractality, all of which provide a measure of the degree or strength of multifractality exhibited by the sample of data. Crucially however, the strength of multifractality is measured non-parametrically, 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 with critical values that cannot be universally tabulated. It is demonstrated however that appropriate model-based bootstrap resampling schemes can be easily implemented to approximate the distributions of the test statistics under the null hypothesis of unifractality.

The proposed testing methodology is applied to simulated unifractal and multifractal data in a series of Monte Carlo exercises, where it is found to have good empirical size and power properties in wide range of situations and to significantly outperform the existing wavelet-based tests of Wendt and Abry (2007). The power of the tests is shown to be robust against various forms of multifractality under the alternative. Furthermore, the tests perform well for sample sizes that would be considered small in the multifractality literature, thus confirming the suitability of the methodology for the study of either local or global scaling properties. The ability to test the local scaling properties of a process around a particular point in time is particularly beneficial in financial applications, given the recent interest in the class of multifractional processes in work such as Bianchi et al. (2013) and Corlay et al. (2014) as an alternative to multifractal processes for modelling return series.

Finally, the testing methodology is applied in an empirical exercise to study the local scaling properties of a dataset containing 5-minute intraday returns for major exchange rate and equity index series. Although we firmly reject the null hypothesis of globally unifractal distributional scaling for the entire return series available for all assets, when applying our tests to short subsamples of returns we find some evidence of locally unifractal behaviour. Whilst more detailed analysis is required, we argue that this may tentatively be interpreted as evidence in support of multifractional rather than multifractal behaviour in asset returns, suggesting that the use of such processes for financial applications should be further investigated.

The structure of the paper is as follows: Section 2 briefly reviews the relevant theoretical properties of unifractal and multifractal processes and discusses the informal graphical testing

approach previously employed in financial applications. Section 3 presents the set of statistics proposed for testing the null of unifractality versus multifractality and the asymptotic theory of the tests under the null hypothesis of unifractality. Section 4 presents a model-based bootstrap resampling scheme proposed for obtaining the distributions of these statistics. Section 5 contains a Monte Carlo exercise to assess the finite-sample properties of the tests with respect to size and power. Section 6 presents an empirical application of the methodology to exchange rate data. Section 7 concludes. Technical content discussing estimation of the multifractal detrended fluctuation analysis estimator and simulation of multifractal random walk processes in discrete time are found in Appendices A and B, respectively.

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 2.1 begins with a brief review of the relevant theoretical properties of these processes², followed in Section 2.2 by a discussion of the informal graphical testing methodology previously used in the literature to distinguish between unifractal and multifractal scaling and the potential problems with this approach.

2.1 Theoretical Scaling Properties

One way in which multifractal processes and the simpler class of unifractal processes can be defined is in terms of their distributional scaling or scale invariance. Intuitively, 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.

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 $X_{t+\Delta t} - X_t$ is unifractal or multifractal if these increments are stationary³ and satisfy:

$$E[|X_{t+\Delta t} - X_t|^q] \propto \Delta t^{\tau(q)+1} \quad \text{or equivalently} \quad E[|X_{t+\Delta t} - X_t|^q] \propto \Delta t^{qH(q)} \quad (2.1)$$

²More detailed treatments of these topics can be found in Mandelbrot et al. (1997), Calvet and Fisher (2002) or Kantelhardt (2009).

³The stationarity requirement for the increments does not have to hold for the original time series and unifractal or multifractal time series may be non-stationary. However, in this case an appropriate transformation must first be applied that results in stationary increments, before equation (2.1) will hold. The majority of estimators for $H(q)$ and $\tau(q)$ employed in the literature, including those used in the current work, automatically perform local detrending to remove non-stationarities, which is typically sufficient for the case of financial returns.

for all sampling intervals or timescales, Δt , and all $q \in \mathbb{R}$. The functions $\tau(q)$ and $H(q)$ are known as the *scaling function* and *generalised Hurst exponent* respectively. Both representations in equation (2.1) are employed in the literature, but are completely equivalent, with the generalised Hurst exponent and scaling function related via the identity $\tau(q) \equiv qH(q) - 1$.

In either case, the function $H(q)$ or $\tau(q)$ characterises 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 (2.1) describes how the moments of the absolute returns scale with changes in the return timescale Δt . Numerous empirical studies, such as those cited in the Introduction, have confirmed the presence of distributional scaling consistent with equation (2.1) in the return series for a wide range of financial assets.

The distributional scaling law in equation (2.1) holds for both multifractal and unifractal processes, however the functional forms for $H(q)$ and $\tau(q)$ differ between the two classes of process, with those of multifractal processes being more general and thus allowing for more flexible scaling properties. Although other formal definitions of unifractal and multifractal processes are employed elsewhere in the mathematics and physics literature, in the current work we rely on the following definition, which is commonly employed in the finance literature (see for example Calvet and Fisher, 2002):

Definition 1: *We define a multifractal process as a process for which the scaling function $\tau(q)$ of the increment process is a strictly concave function of q and a unifractal process as one for which $\tau(q)$ is a linear function of q . Equivalently, for a multifractal process the generalised Hurst exponent $H(q)$ is a strictly decreasing function of q and for a unifractal process $H(q)$ is constant and independent of q , with this constant value denoted by H .*

Although the conditions in Definition 1 clearly do not define unique functional forms for $H(q)$ or $\tau(q)$ for the class of multifractal processes as a whole, for specific theoretical multifractal processes it is often possible to derive closed-form expressions for $\tau(q)$ and $H(q)$ as functions of the model parameters. For example, the scaling function for the multifractal random walk (MRW) process of Bacry et al. (2001) and Muzy et al. (2001) is completely characterised by the parameter $\lambda > 0$, via $\tau(q) = -(\lambda^2/2)q^2 + (1/2 + \lambda^2)q - 1$; the limiting value of $\lambda = 0$ results in a linear unifractal scaling function and thus λ controls the strength of multifractality for the process. Figure 1 plots the theoretical $\tau(q)$ and $H(q)$ functions for a MRW process with several values of λ^2 in the range typically obtained for financial data in previous empirical studies (see for example Muzy et al., 2001).

From Definition 1, for any unifractal process the generalised Hurst exponent is of the form $H(q) = H$ for all q , and from the identity above relating $H(q)$ and $\tau(q)$ it follows that the

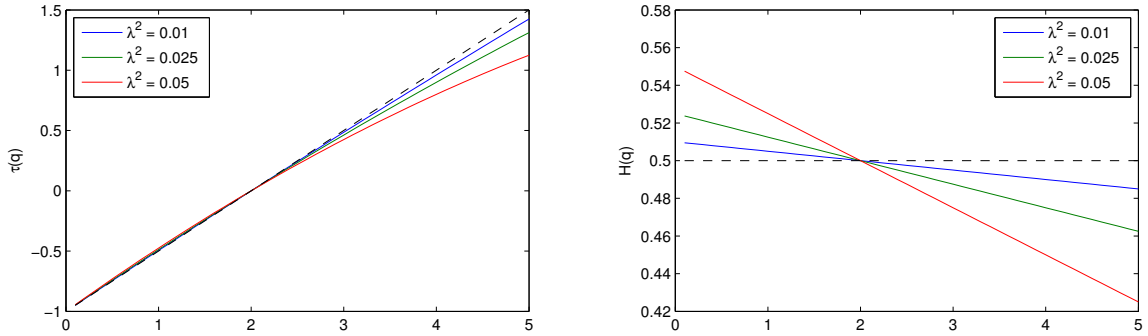


Figure 1: Theoretical values of the scaling function (left) and generalised Hurst exponent (right) for a multifractal random walk process for various values of the parameter λ^2

unifractal scaling function is linear and 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 a common value of H will share the same theoretical scaling properties, with identical functional forms for $H(q)$ and $\tau(q)$. The constant scalar 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. Figure 2 plots the theoretical values of the generalised Hurst exponent and the scaling function against q for a generic unifractal process for various values of the Hurst exponent, H .

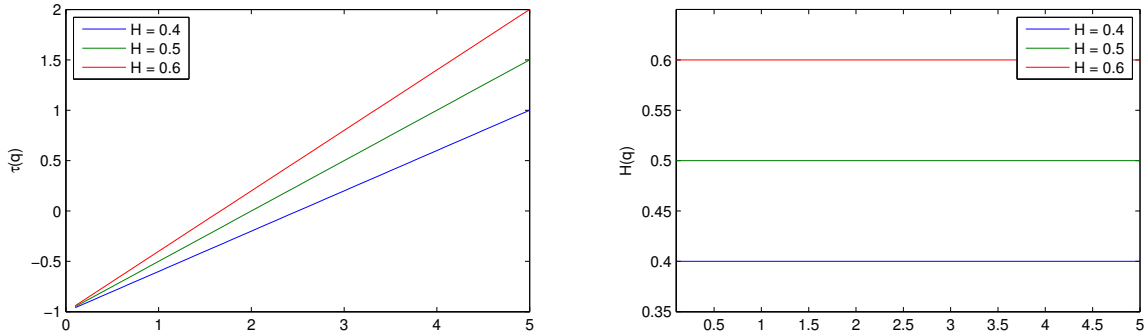


Figure 2: Theoretical values of the scaling function (left) and generalised Hurst exponent (right) for a generic unifractal process for various values of the simple Hurst exponent, H

Whilst simpler than multifractal processes from both a theoretical and practical perspective, pure unifractal processes are unsatisfactory for modelling financial return series, perhaps most notably because they are unable to reproduce the volatility clustering observed empirically in returns. However, a more appropriate class of processes for modelling return series can be obtained simply by allowing the scalar Hurst exponent, H , to be time-varying, with dynamics described by some general continuous or discontinuous function of time, $H(t)$. The resulting class of processes are known as *multifractal* processes, with the function $H(t)$

⁴The term *self-affinity index* is also widely employed in the literature on unifractal processes, however the Hurst exponent terminology will be employed here to remain consistent with the multifractal case.

referred to as the Hurst function. These processes behave locally as unifractal processes, but when viewed globally the time variation in $H(t)$ allows the process to exhibit much more flexible dynamic properties.

Despite the fact that multifractal processes have attracted interest in the finance literature for some time, it is only more recently that the financial applications of multifractal processes have been studied; work such as Frezza (2012), Bianchi et al. (2013), Corlay et al. (2014) and Hallam and Olmo (2014) has demonstrated that with appropriate specifications of the function $H(t)$ (which may be a continuous or discontinuous function of time), theoretical multifractal processes such as the multifractal Brownian motion are capable of reproducing the key properties of return series observed empirically including volatility clustering⁵ and more complex forms of distributional scaling. Therefore, despite the assumption of a globally unifractal return series being somewhat implausible, the assumption of local unifractality implied by a multifractal return process may provide a reasonable alternative to the class of multifractal processes.

Whilst multifractal and multifractional processes both represent generalisations of the class of unifractal processes⁶ that are potentially appropriate for modelling return series, it is important to note that these two generalisations are distinct, in the sense that neither nests the other as a special case; for the former the scaling exponent is constant over time but varies with the moment order, q , and for the latter the situation is reversed. Therefore, the appropriate theoretical models and associated methods for parameter estimation, simulation and other problems differ for these two classes of process, making it important to be able to distinguish between multifractal and multifractional data generating processes.

2.2 Previous Graphical Tests for Distributional Scaling & the Problem of Spurious Multifractality

From Definition 1, strict concavity of the scaling function $\tau(q)$ or a strictly decreasing generalised Hurst exponent $H(q)$ implies a multifractal process and conversely linearity of $\tau(q)$ or a constant $H(q)$ for all q implies a unifractal process. Previous empirical studies have employed informal graphical methods based on this principle to test for the presence and type of distributional scaling in financial return data (see for example Calvet and Fisher, 2002 or Di Matteo, 2007).

This existing graphical testing approach is implemented by estimating the scaling function⁷ $\tau(q)$ for a given sample of data and then visually inspecting it for concavity or linearity.

⁵This is possible because for the majority of theoretical unifractal processes, the parameter H controls not only the distributional scaling properties, but also the smoothness and volatility of the process.

⁶In the sense that both multifractional and multifractal processes nest the simplest unifractal class as a special case, when $H(t) = H$ in all time periods or $H(q) = H$ for all moment orders $q \in \mathbb{R}$ respectively.

⁷In principle estimates of the generalised Hurst exponent $H(q)$ could also be used, but in practice previous work exclusively employs the scaling function representation.

Previous empirical studies applying this graphical testing approach to financial data have interpreted any visible concavity of $\tau(q)$, regardless of magnitude, as a confirmation of multifractal distributional scaling in the series of interest. On this basis, the majority of previous empirical studies have found evidence supporting the presence of multifractal distributional scaling for a range of assets, which has been a key justification for the development and use of multifractal processes for financial returns. However, this graphical testing approach has limitations that have not been adequately acknowledged or addressed.

Although 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, the true population values of these functions are not observable in practice and analysis must be based on sample estimates. The formal statistical properties of most common estimators for multifractal scaling properties are complex and have not been derived⁸, however it is straightforward to demonstrate informally that finite-sample estimation error may produce estimates of $\tau(q)$ and $H(q)$ consistent with a multifractal process even when applied to simulated data from a purely unifractal data generating process. 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).

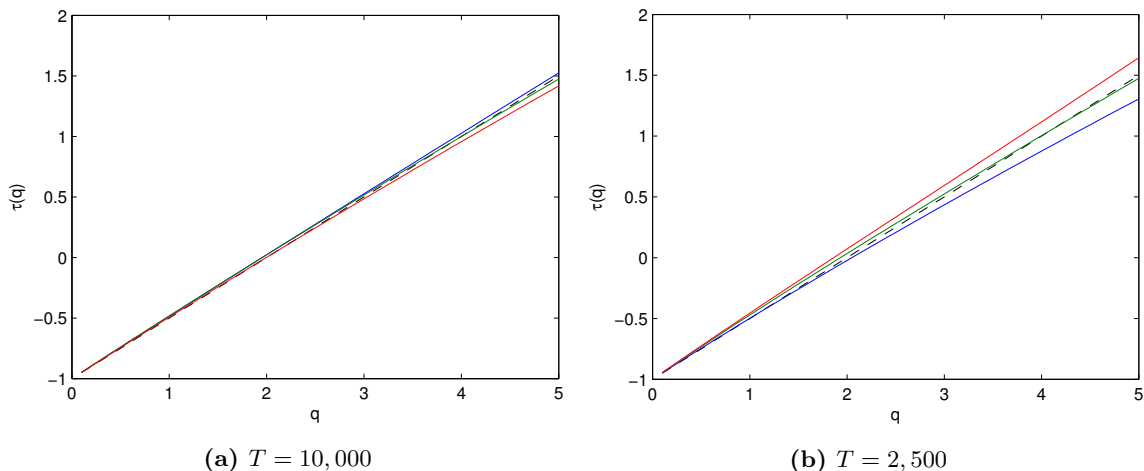


Figure 3: Sample estimates of the scaling function $\tau(q)$ for three realisations from a unifractal fractional Gaussian noise process of length T observations. Dashed lines show the true theoretical values.

As an illustration of this problem, Figure 3 contains the estimated scaling functions for three realisations from a purely unifractal fractional Gaussian noise process of length of $T = 10,000$ and $T = 2,500$, using the MF-DFA estimator of Section 3.1. From Figure 3(a), with $T = 10,000$ the deviation from unifractal scaling is difficult to detect graphically from

⁸See Lashermes et al. (2005) and Schumann and Kantelhardt (2011) for some informal exploration of this topic using simulated data.

the sample estimates of $\tau(q)$, however from Figure 3(b) the estimates of the scaling function for the shorter series length of 2,500 show visible concavity consistent with a multifractal process.

Given that the fundamental cause of this problem is the estimation error for the scaling function or generalised Hurst exponent for finite samples of data, this problem typically becomes less severe as the sample size increases. In principle therefore it can be minimised by using a larger sample of data, however as demonstrated by Figure 3 the problem may persist in samples of several thousand observations and thus there are situations encountered in financial applications where obtaining a sufficiently large sample may not be possible. For example if high-frequency data are unavailable, as in the majority of existing empirical studies, then a sufficiently large sample of data can only be obtained by considering a very long sample period, in which case the observations from the beginning of the sample period may not be informative about the current properties of the process.

Crucially for financial applications, this problem becomes even more pronounced if the true data generating process is multifractional rather than unifractal, which as discussed above is a more realistic alternative class of process for return series than a globally unifractal process. In this case, the time variation in the Hurst function, $H(t)$, typically makes the process appear to be multifractal when viewed over longer time periods, but is revealed to be locally unifractal over shorter sub-periods (see in particular Bianchi and Pianese, 2007 and Schumann and Kantelhardt, 2011). Unlike the case of spurious multifractality associated with a true globally unifractal process, spurious multifractality caused by a multifractional data generating process may not disappear even as the sample size is increased. Spurious multifractality caused by multifractional scaling could in principle be distinguished from true multifractal scaling by testing the local scaling properties of the process using short sub-periods of data; a true multifractal process should be found to be both locally and globally multifractal, whereas a multifractional process will falsely appear globally multifractal, but will be revealed to be locally unifractal when studied over sufficiently short time periods. However, the need for large sample sizes when using the existing graphical testing approach renders it unsuitable for this task. Clearly therefore a more formal testing approach is required that incorporates the effects of estimation error and can be employed in small samples to distinguish between multifractal and unifractal distributional scaling, both at a global and a local level.

A related problem not widely known or discussed in the financial side of the literature on fractal processes is the so-called linearisation effect first formally studied by Molchan (1996) (see also Abry et al., 2007). This effect, arising through the dependence structure of some multifractal processes and their heavy tailed distributions, can lead the estimated sample scaling functions to display linearity for values of the moment order q larger than some threshold value. As such, the effect can result in sample scaling functions that indicate

unifractal behaviour for multifractal processes, with the threshold for q beyond which this value begins to occur potentially being as low as 2-3 for some multifractal processes.

3 Testing Methodology

The current section proposes several possible test statistics that can be used to assess whether the data are most consistent with a unifractal or multifractal process. Section 3.1 presents the various test statistics proposed in this paper for distinguishing between unifractal and multifractal scaling behaviour, and Section 3.2 derives some key results concerning the asymptotic theory of the proposed test statistics. For completeness, Appendix A introduces the multifractal detrended fluctuation analysis estimator employed to estimate the distributional scaling properties of the sample of interest.

3.1 Hypothesis Tests of Unifractality Versus Multifractality

Whilst the graphical tests for unifractality versus multifractality discussed in Section 2 are clearly flawed, the problems arise from the informal nature of the testing approach and their failure to account for the effects of estimation uncertainty, rather than the underlying theoretical basis of the testing method. Indeed, the differences in functional form of $H(q)$ or $\tau(q)$ that exist between unifractal and multifractal processes provide the most natural method for distinguishing between the two classes of process and so will also be used as a basis for the testing procedure proposed here. The problem therefore is to formally test whether the deviation from unifractal behaviour exhibited by the sample estimates of $H(q)$ or $\tau(q)$ is sufficiently large that it cannot be attributed solely to the estimation error.

If viewing the testing problem from a traditional econometric modelling perspective, a suitable test statistic could be constructed by 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. A test could then be 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 choose a functional form for $H(q)$ or $\tau(q)$ under the alternative of multifractality; in the multifractal case, the functional forms differ from one process to another and so choosing a specific functional form makes the test one of unifractality versus a specific multifractal specification, rather than a general test of unifractality versus multifractality⁹.

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

⁹Despite this theoretical limitation, such a testing approach was nonetheless briefly explored, but was also found to perform poorly in practice and so was not pursued further.

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 permits a general test of unifractality versus multifractality, without the need to assume a specific multifractal process under the alternative.

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 $H'(q) = 0$ for all $q \in \mathbb{R}$. 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 $\Delta q \equiv q_2 - q_1$. The corresponding estimates of the generalised Hurst exponent for each $q_i \in \mathbf{q}$ are denoted by $\widehat{H}(q_i)$ and are obtained from the estimated generalised Hurst exponent values using the central difference numerical differentiation approach via:

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

where $\widehat{H}'(q_i, \Delta q)$ denotes the approximate derivative of $H(q)$ at q_i , with an interval of length Δq between each of the points.

Given the condition $H'(q) = 0$ for all $q \in \mathbb{R}$ for any unifractal process, there are many possible functions of $\widehat{H}'(q_i, \Delta q)$ that could be used as statistics for testing the null hypothesis of unifractality, depending on the definition of the alternative hypothesis. One possibility is given by the infimum of $\widehat{H}'(q, \Delta q)$ taken over all $q \in \Omega$, with $\Omega \subset \mathbb{R}$ a compact set. The corresponding test statistic is:

$$d\widehat{H}_{\text{inf}} = \sqrt{N_s} \inf_{i=2, \dots, k-1} \left[\widehat{H}'(q_i, \Delta q) \right], \quad (3.1)$$

with N_s as defined in Section 3.1 above. For a true unifractal process, the population value of the $d\widehat{H}_{\text{inf}}$ statistic in equation (3.1), given by $dH_{\text{inf}} = \inf_{q \in \Omega} [H'(q)]$, 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 $H_0 : dH_{\text{inf}} = 0$ vs $H_1 : dH_{\text{inf}} < 0$.

In principle a closely related alternative test statistic could be obtained by replacing the infimum of $\widehat{H}'(q_i, \Delta q)$ in equation (3.1) with the supremum to give:

$$d\widehat{H}_{\text{sup}} = \sqrt{N_s} \sup_{i=2, \dots, k-1} \left[\widehat{H}'(q_i, \Delta q) \right]$$

where the null hypothesis of unifractality again corresponds to $H_0 : dH_{\text{sup}} = 0$, which is tested against the alternative $H_1 : dH_{\text{sup}} < 0$ with $dH_{\text{sup}} = \sup_{q \in \Omega} [H'(q, \Delta q)]$. Whilst both null hypotheses $H_0 : dH_{\text{inf}} = 0$ and $H_0 : dH_{\text{sup}} = 0$ correspond to unifractality, the

alternative hypotheses of these two tests are subtly different; in both cases, rejection occurs for negative sample values, so that only deviations consistent with a multifractal process (i.e. $H'(q, \Delta q) < 0$) are considered, however $H_1 : dH_{\text{inf}} < 0$ implies multifractal scaling for *some* $q \in \Omega$, whereas $H_1 : dH_{\text{sup}} < 0$ implies multifractal scaling for *all* $q \in \Omega$. Because of this distinction, the alternative $H_1 : dH_{\text{sup}} < 0$ corresponds specifically to multifractality, whereas $H_1 : dH_{\text{inf}} < 0$ corresponds to no unifractality, which includes multifractality as a possibility but does not guarantee it.

This distinction may prove advantageous for the dH_{sup} statistic if true data generating process may be neither unifractal nor multifractal, but if it is known (or can be assumed) that the true process is either unifractal or multifractal, then the dH_{inf} statistic is likely to perform better in practice. This is because the value of the dH_{sup} statistic is determined by the smallest deviation from unifractal scaling for $q \in \Omega$, whereas that of the dH_{inf} statistic is determined by the largest deviation¹⁰. The latter is likely to be more informative about the true scaling properties of the data than the former, since even estimates of $H(q)$ for purely unifractal processes will typically exhibit small deviations consistent with multifractal scaling for some values of q due to estimation error. This was confirmed in practice by initially including both the $d\hat{H}_{\text{inf}}$ and $d\hat{H}_{\text{sup}}$ statistics in the Monte Carlo exercises and comparing their performance; in larger samples the differences were 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. Therefore, the alternative dH_{sup} statistic was omitted from the final Monte Carlo exercise of Section 5 in favour of the dH_{inf} statistic.

The final test statistic based on $H(q)$ considered here is based on the average deviation from unifractal distributional scaling and is constructed as:

$$d\hat{H}_{\text{avg}} = \sqrt{N_s} \left(\frac{1}{k-2} \sum_{i=2}^{k-1} \left| \hat{H}'(q_i, \Delta q) \right| \right) \quad (3.2)$$

This $d\hat{H}_{\text{avg}}$ statistic measures the average deviation in absolute value from unifractal scaling of $\hat{H}(q)$ over the chosen domain of q . Again, for a unifractal process the value of the true population analogue of equation (3.2), $dH_{\text{avg}} = E_q[|H'(q, \Delta q)|]$ equals zero, where $E_q[\cdot]$ is the expected value defined over $q \in \Omega$. Note that for the $d\hat{H}_{\text{avg}}$ 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 offset each other and produce a value of the test statistic 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

¹⁰Note again that in both cases rejection only occurs for negative sample values, so that only deviations consistent with a multifractal process for a given q are considered.

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 $d\widehat{H}_{\text{inf}}$ statistic above, the alternative of $H_1 : dH_{\text{avg}} > 0$ technically corresponds to no unifractality, rather than multifractality, but a test for unifractality versus multifractality can be implemented as discussed above.

Test statistics based on 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, statistics based on both functions have been included to demonstrate how test statistics can be formed in each case and to explore in the Monte Carlo exercise of Section 5 whether there are any advantages in practice of basing tests on either representation of scaling behaviour.

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 $\tau''(q) = 0$ for all $q \in \mathbb{R}$. 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 $\tau''(q) < 0$ for all $q \in \Omega \subset \mathbb{R}$. 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 $\widehat{\tau}(q_i)$ for $i = 1, \dots, k$. An estimate of $\tau''(q)$ at $q = q_i$ can then be obtained by applying numerical differentiation methods to the estimates $\widehat{\tau}(q_i)$, via:

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

As with the previous test statistics based on $H'(q)$, various functions of $\widehat{\tau}''(q, \Delta q)$ could be employed as test statistics in the current context, however attention is restricted to direct analogues of the earlier statistics based on $\widehat{H}'(q, \Delta q)$ in equations (3.1) and (3.2):

$$d\widehat{\tau}_{\text{inf}} = \sqrt{N_s} \inf_{i=2, \dots, k-2} [\widehat{\tau}''(q_i, \Delta q)] \quad \text{or} \quad d\widehat{\tau}_{\text{avg}} = \sqrt{N_s} \left(\frac{1}{k-2} \sum_{i=1}^{k-2} |\widehat{\tau}''(q_i, \Delta q)| \right) \quad (3.3)$$

The interpretation of the $d\widehat{\tau}_{\text{inf}}$ and $d\widehat{\tau}_{\text{avg}}$ test statistics above is identical to that of the earlier test statistics; the $d\widehat{\tau}_{\text{inf}}$ statistic measures the largest deviation from unifractal scaling behaviour over the chosen domain of q , whereas the $d\widehat{\tau}_{\text{avg}}$ statistic measures the average absolute deviation from unifractal scaling over the same domain. The linear scaling function of a unifractal process implies that the true population values $d\tau_{\text{inf}}$ and $d\tau_{\text{avg}}$ given by:

$$d\tau_{\text{inf}} = \inf_{q \in \Omega} [\tau''(q, \Delta q)] \quad \text{and} \quad d\tau_{\text{avg}} = E_q[|\tau''(q, \Delta q)|]$$

should be equal to zero, whereas the strict concavity of the scaling function for a multifractal

process implies $d\tau_{\text{inf}} < 0$ and $d\tau_{\text{avg}} > 0$. A test of the null hypothesis of unifractality against the alternative of multifractality¹¹ 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 $H_1 : d\tau_{\text{avg}} > 0$.

Our test statistics based on the scaling function, $\tau(q)$, can also be connected with an alternative measure for the strength of multifractality sometimes used in the literature. The so-called intermittency coefficient, typically denoted by λ^2 and defined as $\tau''(0)$, provides a single scalar measure of the strength of multifractality, with larger values of λ^2 indicating stronger multifractality and $\lambda^2 = 0$ corresponding to unifractality. Our test statistics based on $\tau(q)$ can thus be viewed as a generalisation of the intermittency coefficient that consider the curvature of the scaling function for a range of values, rather than at a single point.

Alternatively, but more ambitiously, a testing approach could also be based not on the moment scaling properties of the process, but instead on characteristics of the serial dependence structure of multifractal processes. If we denote by $r(t, \tau)$ the continuous return at time t for a return interval of τ , then it can be shown (see e.g. Muzy et al., 2001 or Bacry et al., 2008) that the covariance of the logarithmic absolute returns, $\omega(t, \tau) \equiv \ln |r(t, \tau)|$ for a multifractal process satisfy the following relationship:

$$\text{Cov}[\omega(t + \Delta t, \tau), \omega(t, \tau)] \simeq \lambda^2 \ln \left(\frac{T}{\Delta t} \right)$$

where λ^2 is the previously discussed intermittency coefficient. The return interval τ can be fixed at the finest scale of 1 and a test for the null of unifractality corresponds to the $\omega(t, 1)$ series being short-range correlated (since for a unifractal process $\lambda^2 = 0$). Such a testing approach follows from the arguments of Muzy et al. (2001) and others, however an implementable test based on this concept has not to our knowledge been developed. Such a test could potentially be implemented as a form of Portmanteau test, however in the current work we focus on our approach based on moment scaling analysis. This is largely to maintain consistency with the existing literature on graphical testing methods, although this alternative testing approach based on the serial correlation structure could be pursued as an avenue of future research.

3.2 Asymptotic Theory

This section develops the asymptotic theory for the family of test statistics proposed above under the null hypothesis of unifractality. The theoretical results derived here build on earlier work of Bardet and Kammoun (2008), who previously developed the asymptotic properties for the DFA estimator for the simple Hurst exponent H of a unifractal process. We now

¹¹As 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.

extend these results to the MF-DFA estimator of $H(q)$ for different values of q , when this estimator is applied to a *unifractal* process under the null. More specifically, under the null hypothesis of unifractality, it follows that:

$$E[|X_{t+s} - X_t|^q] \propto s^{qH}. \quad (3.4)$$

The empirical counterpart of the expectation term is equal to $F_{s,q}^q$, which under the null hypothesis of unifractality satisfies $F_{s,q}^q \propto s^{qH}$, for $s > 1$.

Bardet and Kammoun (2008) develop theoretical results for the DFA estimator under the assumption of a specific unifractal process, namely the fractional Gaussian noise (see Appendix A for additional details of this process). Then, relying on previous theoretical results, they argue that because a wide range of unifractal processes converge in law to a fractional Gaussian noise, the theoretical results obtained for this specific process also apply asymptotically for a broad class of unifractal processes (see the original work for further details). We follow an equivalent approach here, with the Hurst exponent and standard deviation of the underlying fractional Gaussian noise denoted by H and σ^2 respectively.

Let $\bar{s} \in \{1, \dots, T\}$, $N_s = \lfloor T/\bar{s} \rfloor$ and $r_i \in \{1, \dots, N_s\}$ for each $i = 1, \dots, m$, with $r_1 < \dots < r_m$. The m window sizes s_1, \dots, s_m are then given by $s_i = r_i \bar{s}$. A necessary condition on \bar{s} to derive appropriate asymptotic theory is that $T/\bar{s} \rightarrow \infty$ and $T/\bar{s}^3 \rightarrow 0$ when $T \rightarrow \infty$. Define $\widehat{S}_{\nu,q} = \frac{F_{\nu,s}^q(s)}{s^{qH}\sigma^2 f(H)}$ and $\widehat{S}_{s,q} = \frac{F_{s,q}^q}{s^{qH}\sigma^2 f(H)}$ with $f(H) = \frac{1-H}{(2H+1)(H+1)(H+2)}$ as derived in Bardet and Kammoun (2008). Note that $\widehat{S}_{s,q}$ can be written in terms of a sum of the random variables $(\widehat{S}_{\nu,q})_{1 \leq \nu \leq \lfloor N_s \rfloor}$. Further, $\widehat{S}_{\nu,q} = \widehat{S}_{\nu}^{q/2}$ with $\widehat{S}_{\nu} = \frac{F_{\nu,s}^2}{s^{2H}\sigma^2 f(H)}$ defined in Bardet and Kammoun (2008). These authors derive the first statistical moments of this sequence of random variables and prove a central limit theorem satisfied by the logarithm of the empirical mean of such sequence. In what follows, we use a similar strategy to derive the asymptotic distribution of $\widehat{S}_{s,q}$. More specifically:

$$\widehat{S}_{s,q} = \frac{1}{s^{qH+1}\sigma^2 f(H)} \sum_{\nu=1}^{N_s s} (F_{\nu,s}^2)^{q/2} = \frac{1}{s^{qH+1}\sigma^2 f(H)} \sum_{\nu=1}^{N_s s} Z_{\nu}^q$$

where the zero-mean Gaussian vector $Z = (Z_1, \dots, Z_T)$ has the covariance matrix $P\Sigma P$, where P is a diagonal block matrix with each block consisting of an (s, s) matrix P_{E_1} defined as the matrix of orthogonal projections on a vector space E_1 defined by $e_1 = (1, \dots, 1)$ and $e_2 = (1, 2, \dots, s)$; Σ is the covariance matrix of the fractional Brownian motion time series. Using a Lindeberg condition, $\widehat{S}_{s,q}$ satisfies the following central limit theorem:

$$\sqrt{N_s} \left(\widehat{S}_{s,q} - E[\widehat{S}_{\nu,q}] \right) \rightarrow N(0, \widetilde{\gamma}^2(H)), \quad (3.5)$$

with $\widetilde{\gamma}^2(H)$ the asymptotic variance of $\widehat{S}_{\nu,q}$, if $\lambda = \|P\Sigma P\|$, the supremum of the eigenvalues

of the symmetrical matrix $P\Sigma P$, is such that $\lambda = o\left(\frac{1}{\sqrt{N_s}}\right)$. This property can be proved applying Lemma 4.1 and Step 2 in the proof of Proposition 3.2 in Bardet and Kammoun (2008). Applying the delta method:

$$\sqrt{N_s} \log \widehat{S}_{s,q} \rightarrow N(0, \widetilde{\gamma}^2(H)). \quad (3.6)$$

This result can be obtained for different lengths of windows satisfying the conditions $N_s \rightarrow \infty$ and $T/\bar{s}^3 \rightarrow 0$. Let (s_1, \dots, s_m) be such different window lengths. Then, one can write for T and s_i large enough:

$$\log \widehat{S}_{s,q} \approx \frac{1}{\sqrt{[N_s]}} \varepsilon_i \Rightarrow \log F_{s_i,q} \approx H \log(s_i) + \frac{1}{2} \log(\sigma^2 f(H)) + \frac{1}{\sqrt{[N_s]}} \varepsilon_i,$$

with $\varepsilon_i \sim N(0, \widetilde{\gamma}^2(H))$. Note that for each q , the MF-DFA estimator is obtained through the OLS estimation of H . Thus:

$$\widehat{H}(q) = \frac{\sum_{i=1}^m (\log F(s_i) - \overline{\log F(s_i)})(\log s_i - \overline{\log s_i})}{\sum_{i=1}^m (\log s_i - \overline{\log s_i})^2}$$

that under some algebra is shown to satisfy $\widehat{H}(q) = H + \frac{1}{\sqrt{N_s}} \varepsilon_i$. Then, it is immediate that $\sqrt{N_s}(\widehat{H}(q) - H)$ converges in distribution to a Normal distribution. Similarly, under the null hypothesis, we can prove that for each q , $\sqrt{N_s} \widehat{H}'(q, \Delta q)$ converges to a Normal distribution and the vector $(\widehat{H}'(q_1, \Delta q_1), \dots, \widehat{H}'(q_k, \Delta q_k))$ to a multivariate Normal (MN) distribution. More formally:

$$\sqrt{N_s} (\widehat{H}'(q_1, \Delta q_1), \dots, \widehat{H}'(q_k, \Delta q_k)) \rightarrow MN(0, \widetilde{\Sigma}(H)), \quad (3.7)$$

with $\widetilde{\Sigma}(H)$ the corresponding asymptotic covariance matrix.

The tightness of the process $\widehat{H}'(q, \Delta q)$ on $q \in \Omega$ implies that the above result can be extended to the functional space $l^1(\infty)$. Then, under the null hypothesis of unifractality, the process $\widehat{H}'(q, \Delta q)$ satisfies:

$$\sqrt{N_s} \widehat{H}'(q, \Delta q) \rightarrow G_\infty(q; H), \quad (3.8)$$

with $G_\infty(q; H)$ a zero-mean Gaussian process with covariance function given by the functional counterpart of the covariance matrix $\widetilde{\Sigma}(H)$. Finally, the convergence of the different functionals of $\widehat{H}'(q)$ used as test statistics in the preceding subsection follows from the continuous mapping theorem applied to the functional space.

4 Model-based Bootstrap Resampling

The distributions of the test statistics under the null hypothesis of unifractality are non-standard, due to the presence of nuisance parameters in the asymptotic null distribution. These nuisance parameters are given by functions of the parameter H , which is unknown and must be estimated. Consequently, critical values of the different tests cannot be universally tabulated, however the finite-sample distribution of the test can be approximated using alternative methods explored in the literature, namely, bootstrap and simulation techniques.

Both unifractal and multifractal processes will generally exhibit serial dependence that may persist at long horizons and so the method employed must account for the dependence structure present in the original data. This issue is particularly critical, 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 method that cannot preserve the serial dependence structure of the data will also be unable to preserve their scaling properties and so is likely to lead to invalid inference.

Given the above constraints, the most obvious choice of methods would appear to be non-parametric resampling approaches for dependent data, such as block bootstrap or subsampling techniques. However, as argued below, the use of nonparametric resampling methods is not strictly necessary in the current testing environment and a semi-parametric model-based bootstrap approach may be employed instead. This type of bootstrap takes into account the serial dependence in the data by modelling parametrically the dynamics of the process. For completeness, the performance of both block bootstrap and subsampling methods was also tested in Monte Carlo exercises to investigate whether they could provide superior performance to the model-based bootstrap approach. Unfortunately, for the sample sizes of interest in the current application, these nonparametric approaches were found to perform substantially worse than the model-based bootstrap approach and so were deemed unsuitable for the current context.

In response to the requirements discussed above, a resampling approach was developed for obtaining the null distributions of the test statistics that can be viewed as a model-based bootstrap approach, combining aspects of both Monte Carlo simulation and bootstrap methods. Despite being developed independently for the current work, the resampling scheme proposed here is closely related to the previous model-based bootstrap approach of Hall et al. (2000) proposed for performing statistical inference for the scalar simple Hurst exponent of a unifractal process. The testing problem considered in the current work is however more complex, because we perform inference based on test statistics that are functionals of the functions $H(q)$ and $\tau(q)$ for the more general multifractal case, instead of simply performing inference for the scalar parameter H in the unifractal case. As a result, the resampling algorithm developed for the current context necessarily differs from that proposed by Hall

et al. (2000), although the two approaches share the same underlying theoretical motivation.

Although the resampling approach proposed here is potentially applicable in other testing situations, some important characteristics specific to the current testing environment must be emphasised in order to motivate the use of this method in the current context. Firstly, all of the proposed test statistics are based solely on the scaling properties of the process characterised by 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 all of the proposed test statistics correspond to that of a unifractal process and from Definition 1, the functional forms of the generalised Hurst exponent or scaling function are identical for all unifractal processes and crucially are parameterised entirely by the simple Hurst exponent, H .

Whilst the true scaling properties are identical for any unifractal process with the same simple Hurst exponent, sample estimates of the functions $H(q)$ and $\tau(q)$ must be used in practice, with the resulting estimation error potentially leading to the problem of spurious multifractality. We therefore wish to model this estimation uncertainty 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 unifractal. The discussion above suggests that in order to implement the bootstrap test we need to estimate first the simple Hurst exponent from the original sample of data, under the assumption that the data are consistent with the null hypothesis of unifractality¹².

4.1 Bootstrap Resampling Algorithm

In what follows, we outline the procedure to generate the bootstrap p-values that consistently estimate the asymptotic p-values of the test. First, we simulate a large number of independent sample paths 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, with the resulting estimate denoted by \hat{H} .

¹²At 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 value¹³ \widehat{H} is used to generate R independent realisations from the chosen fractional Gaussian noise data generating process, with each simulated series having the *same length as the original time series*, denoted by T .
3. For each of the R simulated series the generalised Hurst exponent, $H(q)$, or the scaling function, $\tau(q)$, are estimated, with the resulting estimates denoted by $\widetilde{H}(q)_1, \dots, \widetilde{H}(q)_R$ and $\widetilde{\tau}(q)_1, \dots, \widetilde{\tau}(q)_R$. The sample value of the chosen test statistic from Section 3 is then calculated for each of these R estimates, with the sample values for the chosen test statistic denoted generically by $\widetilde{T}_1, \dots, \widetilde{T}_R$ for notational simplicity, rather than for example $d\widetilde{H}_{\text{inf},1}, \dots, d\widetilde{H}_{\text{inf},R}$.
4. Sample estimates of $H(q)$ and $\tau(q)$ are obtained from the original 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 $\widehat{H}(q)$ and $\widehat{\tau}(q)$ and are used to calculate the sample value of the chosen test statistic, denoted by \widehat{T} .
5. The empirical cumulative distribution function of the R bootstrap test statistic values $\widetilde{T}_1, \dots, \widetilde{T}_R$ denoted by $\widetilde{F}(T)$ is calculated as:

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

where $I(\widetilde{T}_i \leq T)$ is the indicator function taking the value of 1 if $\widetilde{T}_i \leq T$ and 0 otherwise. The empirical distribution $\widetilde{F}(T)$ 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 \widehat{T} can be calculated as $\widetilde{F}(\widehat{T})$, with the null of unifractality rejected if the sample p-value is less than the chosen nominal significance level. The average-based statistics are one-sided with rejection regions on the right side of the distribution and thus sample p-values can be calculated as $1 - \widetilde{F}(\widehat{T})$. Alternatively, in both cases the empirical inverse cumulative distribution function can be calculated and used to obtain critical values.

Estimation of the generalised Hurst exponent and scaling function in Steps 3 and 4 is performed using the MF-DFA estimator discussed in Appendix A and estimation of the simple Hurst exponent, H , in Step 1 is performed using the (unifractal) DFA estimator, which is obtained as a special case of the MF-DFA estimator for a value of $q = 2$.

¹³Note that the fractional Gaussian noise process is generally characterised by two parameters; H and the increment standard deviation σ . The latter however is not relevant in the current testing environment, since changes to σ leave the scaling properties unchanged, and can thus be set to any arbitrary value such as $\sigma = 1$.

It should be noted at this point that both the procedure used to estimate the process parameters in step 1 and the fGn DGP used for simulation in step 2 to approximate the null distribution of the test statistic implicitly assume a lack of time variation in the parameters of the process. As a result, our testing procedure does not allow for the presence of structural breaks in the series, though this limitation is of course common to many econometric and statistical tests across the literature.

In addition, the fractional Gaussian noise process employed for our resampling procedure also implies smooth sample paths for the corresponding fractional Gaussian motion price or cumulative sum series and thus technically rules out the possibility of jumps. Nonetheless, additional simulation exercises were performed in order to empirically investigate the magnitude of the impact that jumps would have on test performance by including an additive jump component in the simulation DGPs employed in Section 5 in the manner of Xue et al. (2014) or Laurent et al. (2016). These simulation exercises confirmed that, as expected a priori, test performance is severely affected by the presence of jumps even with relatively small jump magnitudes and low jump frequency.

Therefore, if jumps are suspected to be present in the series of interest we suggest the use of a jump detection approach such as those proposed by Lee and Mykland (2008) or Xue et al. (2014) to identify jumps, before appropriately filtering them out of the raw data. Our testing procedure can then be applied to the continuous or non-jump component of the series. Alternatively, a more substantial modification of the testing procedure may allow it to perform satisfactorily with jumps still present within the estimation sample. Potentially this could be achieved by replacing the DFA estimator with an alternative that is more robust to jumps, together with generalising the fGn DGP used in step 2 of the algorithm to an alternative form of unifractal process does permit jumps. However, both of these changes are likely to substantially reduce the small sample performance of the test relative to our current approach, which was viewed as a priority in the current work.

The next result states the asymptotic validity of the bootstrap approximation. More specifically, under the null hypothesis of unifractal, the process $\tilde{H}'(q, \Delta q)$ obtained from the model-based bootstrap samples introduced in the algorithm above satisfies:

$$\sqrt{N_s} \tilde{H}'(q, \Delta q) \rightarrow \tilde{G}_\infty(q; H), \quad (4.1)$$

with $\tilde{G}_\infty(q; H)$ an independent copy of the zero-mean Gaussian process $G_\infty(q; H)$ introduced in (3.8). The proof of the convergence result (4.1) immediately follows applying the same arguments as for the original series simulated from the the fractional Brownian motion time series discussed above. The weak convergence of the corresponding bootstrap test statistics is derived from applying the continuous mapping theorem to the above process.

4.2 Justification for the Resampling Method

One apparent limitation of the proposed model-based bootstrap resampling algorithm is the need to specify a particular DGP consistent with the null hypothesis to simulate from in Step 3 to estimate the null distribution of the test statistic. Although other theoretical unifractal processes could in principle be employed, we assume that the fractional Gaussian noise (fGn) is used. The fGn is the increment series of the fractional Brownian motion (fBm), which can be viewed as a generalisation of the standard Brownian motion that allows for long-range linear dependence in 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 process¹⁴. For the special case for $H = 0.5$ the fGn is serially uncorrelated and simplifies to the standard Gaussian noise, whereas for $H > 0.5$ and $H < 0.5$ the process displays positive and negative serial correlation respectively. As a Gaussian process, numerous algorithms are available¹⁵ for the simulation of fGn sample paths in Step 2 above, with the well-known circulant embedding method used here.

The first justification for this choice of process is simply that the fGn and the corresponding fBm cumulative sum series are by far the most commonly used unifractal processes in the finance literature and elsewhere, with the relatively small number of suitable alternatives receiving much less attention. The choice of the fGn is still justifiable from this perspective if interest lies in testing the local scaling properties of the process, in which case the null corresponds to a globally multifractional process; the most widely studied theoretical multifractional process is the multifractional Gaussian noise and corresponding multifractional Brownian motion level series, which are obtained as direct generalisations of the fGn or fBm by allowing the simple Hurst exponent to be time-varying. The process thus behaves locally as an fGn and as such, assuming an fGn data generating process under the null of (local) unifractality is still justifiable.

Secondly, existing theoretical results can be employed to argue that the null distribution obtained by simulating from an fGn data generating process is asymptotically identical to the distribution that would be obtained from a wide range of alternative unifractal processes consistent with the null hypothesis. In particular, it has been shown by Taqqu (1975), Giraitis

¹⁴For the fGn the same parameter controls both the scaling and dependence properties of the process. This is a characteristic of the subclass of ‘self-affine’ processes, of which the fBm and fGn are members, however unifractal processes more generally may possess independent scaling and dependence properties. The construction of such generalised unifractal processes is more complex and so few examples exist in the literature, with a notable example being that of Gneiting and Schlather (2004).

¹⁵Indeed a secondary benefit of using the fGn is that the methods available for simulating sample paths include a number of very efficient algorithms, thus reducing computational requirements compared to alternative unifractal processes.

and Surgailis (1985) and others that, under certain regularity conditions, the cumulative sum series for a range of unifractal processes converge in law to a fractional Brownian motion. The chosen MF-DFA estimator (discussed in detail in Appendix A), like most common estimators for the generalised Hurst exponent, operates on the cumulative sum or profile series, implying that the null distribution of the test statistics derived from the estimates of $H(q)$ or $\tau(q)$ for an fGn should be asymptotically identical to that for a much larger class of unifractal processes under the null.

Alternatively, Hall et al. (2000) consider asymptotic equivalence of the fGn and other unifractal processes at the level of the estimator, rather than at the level of the underlying process. In their testing environment, they need only to estimate the simple Hurst exponent, H , for which they employ the rescaled adjusted range (R-S) analysis estimator. The R-S estimator has no multifractal analogue for the generalised Hurst exponent, $H(q)$, and so is inapplicable in the current context. Nonetheless, they demonstrate that under some assumptions and regularity conditions the asymptotic distribution of the R-S estimator of H for an fGn process and that for a wider class of processes obtained via linear and nonlinear transformations of that fGn process, depend only on the value of H and some known nuisance parameters unaffected by the transformation. As a result, the distribution of the R-S estimator of H obtained from a fGn process for given values of H and the other nuisance parameters will be identical to the distribution for a much wider class of processes. Intuitively, this occurs because the R-S estimation algorithm performs standardisation using the local variances that corrects for nonstationarities of various forms, eliminating the effects of the transformation performed for the Gaussian fGn process. As such, the theoretical result derived by Hall et al. (2000) applies specifically to their chosen R-S estimator and cannot be directly applied to the DFA/MF-DFA estimators used in the current context. However, given that the DFA-type estimators also perform a form of local detrending with the specific aim of removing nonstationarities from the data, we hypothesise that a similar property will also hold for the current estimation environment, although we do not prove this formally.

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 a variety of situations. Section 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. Section 5.2 presents a competing test based on wavelet analysis. Section 5.3 reports and discusses the core size and power results, with the robustness of these results to various changes in the testing environment checked in Subsections 5.4 and 5.5.

5.1 Monte Carlo Methodology

A standard Monte Carlo approach is employed for calculating the empirical size and power of the various test statistics under different conditions. In order to calculate the empirical size of the tests under the null of unifractality, we apply the testing methodology to samples of data generated by a unifractal process consistent with the null hypothesis, for which we use the fGn process discussed in the previous section.

As argued in the previous section, a range of other processes consistent with the null of unifractality will be asymptotically equivalent in the current context to the fGn employed for simulation during our model-based bootstrap resampling method, giving our approach greater generality than it first seems. However, the null distribution of the test statistic for such processes in finite samples may differ from that of an fGn process. In this sense, the empirical size figures reported here should therefore be viewed as an upper bound to the performance of the test in terms of its size properties. It would also be beneficial to investigate the finite-sample size properties of the testing procedure under alternative unifractal DGPs, however the widespread use of the fGn and its Gaussian nature have produced simulation algorithms that are significantly more efficient than those for other unifractal processes. Given the significant computational demands of embedding our model-based bootstrap approach within a Monte Carlo exercise, we leave this as an avenue for future investigation.

The outcome of the test at nominal significance levels of 10%, 5% and 1% is recorded for each of the test statistics and 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 applying the testing methodology to simulated data from the multifractal random walk (MRW) process previously mentioned in Section 2. Sample paths of the MRW process are simulated using the discrete time formulation of Bacry et al. (2001) detailed in Appendix B. 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 (the intermittency coefficient), permitting the relationship between power and the strength of multifractality to be easily studied. The Markov-switching multifractal (MSM) also mentioned in Section 2 is employed in Section 5.5 to confirm that the testing procedure has power against an alternative multifractal process appropriate for financial applications.

In all cases both the number of simulated sample paths in stages 2 and 3 of the model-based bootstrap resampling algorithm, R , and the number of Monte Carlo repetitions, M , are set to a value of 1,000. The chosen values of $R = 1,000$ and $M = 1,000$ 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.

The default parameter values for all processes were chosen based on parameter estimates obtained from daily return series for a range of financial assets, with the robustness of the results to different choices for these parameter values explored in Section 5.5. 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 parameter¹⁶. 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.

Finally, given our interest in the local distributional scaling properties of processes, it would be informative to study the properties of the testing procedure when the true DGP is multifractional, such as the multifractional Gaussian noise. In principle such an exercise is implementable, with its current omission being due to practical reasons that would make it substantially more complex and time-consuming than the existing simulation exercises. There has been some work developing methods for estimating the time-varying Hurst function, $H(t)$, under the assumption of a multifractional process for a given sample of data, typically based on a dynamic application of the estimators for the time-invariant Hurst exponent in the unifractal case. However, we are not aware of any approach for generating simulated sample paths of the Hurst function, which would be required in order to simulate sample paths for the corresponding multifractional process itself in the current Monte Carlo context. An appropriate methodology for this task would need to be developed in order to extend the Monte Carlo exercises in this way and so we leave it as an avenue for future research.

Given that the vector of values of q over which the functions are estimated (previously denoted by \mathbf{q} in Appendix A) must be equally spaced for the current test statistics, \mathbf{q} is completely characterised by the minimum and maximum values and the grid or step size between each value, denoted by q_{min} , q_{max} and q_{grid} respectively. Estimates of $H(q)$ obtained for values of q that are large in absolute terms typically become unreliable in smaller 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. As a compromise, the default choices were $q_{min} = 0$, $q_{max} = 2.5$ and $q_{grid} = 0.1$, which

¹⁶The additional parameters for the integral timescale, T , and sampling interval Δ were set to $T = 5,000$ 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.

were found to produce consistently strong performance for the sample sizes considered here. Additional choices for \mathbf{q} were also considered¹⁷, including narrower ranges of $0 \leq q \leq 1.5$ and wider ranges of up to $0 \leq q \leq 5$. Given the previous discussion of the linearisation effect that may occur for larger values of q , it is of particular interest to see if our testing procedure can produce reliable inference using this extended range of $0 \leq q \leq 5$. Therefore, we report results for this wider range of q in Section 5.4, to check the robustness of the testing methodology to the choice of \mathbf{q} .

Finally, the minimum and maximum scale or window sizes, previously denoted s_1 and s_m , were set approximately according to the rules of thumb discussed in Appendix A, with $s_1 = 5$ in all cases and $s_m \approx T/15$, where T is the sample size. Given that inappropriately chosen values for the range of scales can lead to unreliable estimates of scaling properties, we carefully investigated the effects of scale choice on test performance. We found however that for reasonable changes in scales used, test performance displayed little variation. We hypothesise that this is due to the effects of scale choice being automatically taken into account by our model-based bootstrap testing procedure, in which identical scales are used for both the sample and bootstrap data series. Indeed, the same argument can be extended to the previously discussed invariance of test performance to reasonable changes in the range of q employed. In light of this finding, we continued to employ scales roughly consistent with the typical rules of thumb outlined above.

5.2 Wavelet-Based Tests of Distributional Scaling Properties

As previously mentioned in the Introduction, the only comparable statistical test of unifractal versus multifractal distributional scaling is the wavelet-based testing approach of Wendt and Abry (2007) previously proposed in the signal processing literature. In the current work it is thus logical to compare the relative performance of this existing testing approach with that of the new approach proposed in the current work. Unfortunately, direct comparisons are not possible using the original Monte Carlo exercises of Wendt and Abry (2007), because whilst they do select the same theoretical unifractal and multifractal data generating processes for calculating empirical size and power, they use different parameter values and sample sizes that are much less relevant or appropriate for the current financial context than those proposed above. In particular, Wendt and Abry (2007) focus on larger sample sizes, which are less likely to be obtainable for financial data and do not allow the study of local scaling properties. In addition, they consider much stronger levels of multifractality than are plausible for financial return data when calculating empirical power. Both differences serve to improve the reported performance of their test.

Therefore, new Monte Carlo exercises using these wavelet-based tests have been per-

¹⁷Values of q in the range $-5 \leq q \leq 5$ are typical in empirical studies of multifractality in financial data, such as Muzy et al. (2001), Di Matteo (2007) or Bacry et al. (2008).

formed specifically for the current work, to permit a direct comparison with the model-based bootstrap approach proposed here. The wavelet-based testing approach will not be discussed in any detail (the reader is referred to the original work for further information), but it should be noted that Wendt and Abry (2007) propose equivalent testing approaches based on both wavelet leaders and also wavelet coefficients. The former generally produce tests with higher empirical power, but the latter often result in better empirical size properties and so both approaches will be included in the current Monte Carlo exercises. In both cases, the percentile formulation of the relevant test statistic is selected from the various alternatives proposed by Wendt and Abry (2007), given their conclusion that this formulation provided the strongest relative performance. All other aspects of the implementation of the wavelet-based tests follow the approach used for the Monte Carlo exercises in their original paper and the tests are performed using the Matlab code available online that accompanies their paper.

5.3 Core Results & the Effects of Sample Size

Table 1 presents the empirical size and power of the proposed test statistics obtained using the default choices detailed above with sample sizes of 5,000, 2,500 and 1,000. Whilst these appear to be large sample sizes by normal econometric standards, accurate estimation of the generalised Hurst exponent requires a substantial number of observations, with minimum sample sizes in the low thousands normally suggested in the literature and thus making the sample sizes employed here relatively small in reality. 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 1 it can be seen that for the larger sample sizes of $T = 5,000$ and $T = 2,500$ 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 1,000 the empirical sizes do move further from their nominal levels, but 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 0.86 and 0.95 for $T = 5,000$ and dropping to between 0.53 and 0.73 for the smallest sample size of 1,000. For larger sample sizes the power of the tests also appears to be much more consistent across the different nominal significance levels than for $T = 1,000$, however this will also be affected 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 $T = 1,000$ are typically lower than the nominal levels, which will in turn reduce the power of the tests relative to a test performed

Table 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
T = 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
Wavelet coefficients	0.204	0.131	0.038	0.219	0.148	0.057
Wavelet leaders	0.202	0.127	0.041	0.699	0.588	0.392
T = 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
Wavelet coefficients	0.206	0.138	0.047	0.237	0.158	0.058
Wavelet leaders	0.213	0.137	0.051	0.602	0.516	0.313
T = 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
Wavelet coefficients	0.219	0.144	0.054	0.209	0.136	0.058
Wavelet leaders	0.254	0.173	0.061	0.481	0.386	0.176

at the true nominal significance level.

For any of the given sample sizes the relative performance of the four proposed test statistics is largely consistent, suggesting that the approaches based on the generalised Hurst exponent, $H(q)$, and the scaling function, $\tau(q)$, are not only equivalent in theory but also in practice. For the additional Monte Carlo exercises presented in the current section the observed performance differences between the statistics based on $H(q)$ and $\tau(q)$ remained small and so results for the $\tau(q)$ -based test statistics have not been reported for the rest of the section to conserve space, but are available upon request. Tests based on the average and infimum are also found to be approximately equal in terms of empirical size and power. The same was not true however for the supremum-based statistics excluded from the Monte Carlo exercises for the reasons given in Section 3.1, which often produced empirical size values below the nominal significance level, particularly for $T = 1,000$.

Finally, it is clear from the results for the wavelet-based tests of Wendt and Abry (2007) that the new testing approach proposed here provides a substantial improvement in performance over this existing approach. The empirical size of both the wavelet coefficient and wavelet leader tests are consistently higher than the specified nominal significance levels,

with the observed empirical size typically at least 2 or 3 times larger than the specified nominal value. Consistent with the results of the original Monte Carlo exercises performed by Wendt and Abry (2007), the wavelet leader approach is found to attain substantially higher power than the approach based on wavelet coefficients, however the former is still found to achieve significantly lower power than the new test statistics proposed in the current work. Furthermore, it should again be noted that the values for empirical power reported in Table 1 are not adjusted for differences in the size of the tests. Given that the empirical size of the wavelet-based tests is consistently and substantially higher than the specified nominal level, the actual power of the wavelet-based tests for a given significance level is in fact substantially lower than the values listed in Table 1.

5.4 Robustness to Changes in the Estimation Method

Given the importance of the estimation method for $H(q)$ 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 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 is estimated¹⁸. Checking the power of the testing procedure for different ranges of q is also important from the perspective of the linearisation effect discussed previously in Section 2, given that it may lead to spurious findings of unifractality for larger values of q .

Table 2 contains empirical size and power values for various choices of the vector \mathbf{q} . Results are only presented for the sample size of 5,000 in order to conserve space, but similar results were also obtained for the previous sample sizes of 2,500 and 1,000. Table 2(a) reports results using the default values of $q_{min} = 0$ and $q_{max} = 2.5$, but varying q_{grid} from the default value of 0.1 to include the two additional values of 0.05 and 0.25. Table 2(b) employs the same choices for q_{grid} as Table 2(a), but increases the maximum value of q to $q_{max} = 5$, motivated by the choice of $4 \leq q_{max} \leq 5$ being common in previous empirical studies. It should be noted that the previously discussed linearisation effect may manifest itself for threshold values as low as $q = 2$ or $q = 3$. Such a value may therefore potentially fall within the default range used previously for \mathbf{q} , but the extended range spanning up to $q = 5$ should provide a more reliable indication of whether this effect influences test performance.

From the reported values in Table 2 it is clear the performance of the tests remains largely consistent for the various combinations of q_{grid} and q_{max} considered here. Whilst

¹⁸To check robustness to alternative choices of estimator for $H(q)$, the MF-DFA estimator was also substituted for the multifractal centred moving average (MF-CMA) estimator of Schumann and Kantelhardt (2011). The empirical size and power obtained for the testing methodology employing the MF-CMA was very similar to that employing the MF-DFA estimator, with results available upon request.

Table 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
$q_{\text{grid}} = 0.10$						
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
$q_{\text{grid}} = 0.25$						
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
(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
$q_{\text{grid}} = 0.10$						
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
$q_{\text{grid}} = 0.25$						
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

more extreme choices for the vector \mathbf{q} (particularly including values of q that are large in absolute terms) are likely to have an adverse effect on the performance of the testing methodology, the good size and power properties of the tests seem to be preserved across a variety of reasonable choices for \mathbf{q} . This suggests that the effects of the choice of \mathbf{q} on the properties of the estimator for $H(q)$ are being correctly accounted for by the proposed testing methodology and, at least for the DGPs considered here, it is able to correctly account for the effects of spurious multifractality and the linearisation effect discussed previously.

5.5 Choice of Data Generating 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 intermittency coefficient λ^2 , which appears directly as a parameter of the process. As previously discussed, larger values of λ^2 correspond to stronger multifractality and the limiting value of $\lambda^2 = 0$ corresponds 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. Figure 4 plots

the empirical power of the the dH_{inf} and dH_{avg} statistics as a function of the multifractality strength¹⁹ λ^2 for $T = 5,000$ in order to establish whether this predicted convergence holds in practice.

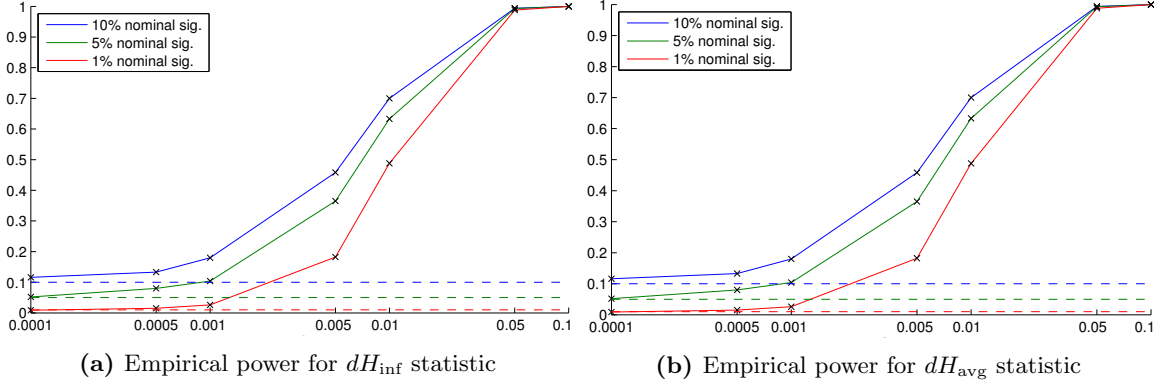


Figure 4: Empirical power versus multifractality strength (λ^2) for dH_{inf} and dH_{avg} statistics. A logarithmic scale has been used for the horizontal axis to improve legibility, but axis values correspond to true values of λ^2 and not logarithmic values.

Figure 4 clearly demonstrates that the expected convergence of test power to the limiting values in both directions is indeed observed empirically. Empirical power of nearly 100% is obtained for all nominal significance levels and test statistics after increasing the strength of multifractality from the default value of $\lambda^2 = 0.025$ to $\lambda^2 = 0.05$ or higher. Moving in the opposite direction, convergence to the nominal significance levels as $\lambda^2 \rightarrow 0$ is almost complete at the value of $\lambda^2 = 0.0001$. It is worth noting from Figure 1(a) that the theoretical scaling function for the MRW process with a value of $\lambda^2 = 0.01$ is sufficiently close to linear that such weak multifractality would be nearly impossible to detect using the previous graphical approach. However, the proposed testing methodology still results in empirical power of approximately 70% and 50% at nominal significance levels of 10% and 1% respectively.

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 developed specifically for financial applications²⁰. We tested two parameterisations of the MSM process, both following from the parameter estimates obtained for financial data in the original work by Calvet and Fisher (2004). For the first parametrisation the number of volatility components \bar{k} was set to 5 and 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

¹⁹Note that for Figure 4 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.

²⁰Additional theoretical multifractal processes have been developed in other fields of study, such as the physics literature. However, given that our interest lies specifically in financial applications we focus on processes developed for this context.

deviation, binomial value, frequency growth rate and high-frequency switching probability respectively (see Calvet and Fisher, 2004, for further details). For the second parametrisation we use 10 volatility components ($\bar{k} = 10$) and values of $\sigma = 0.5, m_0 = 1.3, b = 8, \gamma_{\bar{k}} = 0.975$ for the remaining parameters.

Table 3: Empirical power against Markov-switching multifractal alternative

Nominal sig. level	5 volatility components			10 volatility components		
	0.10	0.05	0.01	0.10	0.05	0.01
T = 5,000						
dH_{inf}	0.995	0.995	0.991	0.974	0.964	0.936
dH_{avg}	0.980	0.977	0.971	0.960	0.947	0.917
T = 2,500						
dH_{inf}	0.985	0.977	0.964	0.936	0.920	0.871
dH_{avg}	0.958	0.950	0.926	0.914	0.885	0.849
T = 1,000						
dH_{inf}	0.924	0.904	0.861	0.833	0.790	0.698
dH_{avg}	0.899	0.879	0.832	0.818	0.776	0.690

Table 3 reports empirical power of the test statistics against these two parameterisations of the MSM, with all other aspects of the testing methodology consistent with the default methodology discussed in Section 4. It is clear from Table 3 that the proposed tests maintain good power 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, direct comparisons of empirical power between these two processes are not possible without a universal measure to compare the strength of multifractality.

Finally, the size and power properties of the test are assessed for series with linear serial dependence. All series simulated thus far have been constructed to exhibit linear independence in the level of the series²¹, based on the empirical finding that financial returns generally exhibit little to no serial correlation in their levels. The fGn process used to calculate empirical size under the null exhibits linear dependence for any $H \neq 0.5$ and the MRW used to measure power under the alternative can be generalised to allow for linear dependence as described in the Appendix. Given that the estimated linear dependence for return series if present is typically weak, 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.

²¹Although 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.

Table 4: 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
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
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

From Table 4 it is clear 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 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 $H = 0.4$, this is not likely to be problematic in practice when applying the testing methodology to real financial data.

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 potential advantages that the proposed testing methodology may have over the existing graphical methods in smaller samples, particular attention is paid to tests of the local, rather than global, scaling properties of the return series.

6.1 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 intervals²².

For the S&P500 and NASDAQ-100 equity data weekends and other non-trading days were removed from the raw price data, with the list of non-weekend closures constructed from the historical list of holidays available on the NYSE website. For the two exchange rate series trading takes place 24 hours a day and 7 days a week, but slows substantially over weekends and certain holidays. Following Andersen et al. (2001), the end of each 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 Year's Day was trading noticeably slower than normal and so only these holidays were omitted from the exchange rate series.

This process leaves a sample size of 1008 trading days for the equity index series, for a total of over 75,000 observations, and 1037 trading days for the exchange rate series, for a sample size of over 250,000. 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).

6.2 Empirical Results

We begin by applying the proposed testing methodology to the complete sample of 5-minute data for each asset to examine the global scaling properties of the return series; for all four financial assets we firmly reject the null of a (globally) unifractal return process, with p-values equal to zero to four decimal places for all of the proposed test statistics. As previously discussed in Section 2.1, a globally unifractal process is overly restrictive and arguably unsuitable as a model for financial returns and so this first empirical finding is largely expected.

However, rejection of the null hypothesis of global unifractality for the complete sample should not immediately be viewed as a confirmation of multifractal distributional scaling in the current dataset, despite the very large samples sizes provided by the use of intraday data. In particular, the class of multifractional processes discussed in Section 2.1 provide a plausible alternative to multifractal processes for modelling return series that locally exhibit

²²This 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).

unifractal distributional scaling, but may falsely appear multifractal even in large samples due to the time variation in the Hurst exponent.

To test for this possibility, we split the complete return series for each asset into consecutive non-overlapping windows of length T and then apply the proposed tests for unifractality versus multifractality to each of these subsamples. Observing evidence of locally unifractal properties within each of these sub-windows, combined with the previous finding of non-unifractal scaling in the full sample, could be interpreted as support for a multifractal process. Whilst this approach does not constitute a formal test for the null of multifractal scaling, which would require a true multiple testing framework, we view it as an informal first step towards such an approach which has value due to the lack of previous work on this topic.

The Hurst function $H(t)$ for a multifractal process may possess any functional form that maps to the interval $[0, 1]$, so the proposed testing approach assumes implicitly that the true underlying function $H(t)$ can be approximated by a step function with a change in value permitted to occur every T observations. By reducing the value of T the quality of this approximation will improve, but very short window lengths are not possible due to the need for relatively large samples for reliable estimation of $H(q)$ or $\tau(q)$. As a result this approach may seem somewhat unsatisfactory, however from a practical perspective it is made more reasonable in the current context due to use of intraday data, which allow large sample sizes to be obtained from relatively short periods of calendar time; for example, windows of 1,000 5-minute observations²³ permit the Hurst exponent to change approximately every 3 trading days for the exchange rate series, thus allowing quite flexible scaling dynamics for the daily return process.

For window lengths of 5,000 or longer, the sample p-values for the null of local unifractality within each sub-period remain equal to zero, again suggesting (locally) multifractal scaling for all four return series. Figures 5 and 6 consider shorter window lengths and plot the sample p-values obtained from applying the proposed test statistics to non-overlapping windows of lengths 2,500 and 1,000 from the 5-minute return series. Note that due to shorter trading hours in equity markets, a smaller number of observations and thus a smaller number of windows are obtained for each window length compared to the exchange rate series.

From panels (a) and (c) of Figure 5, it can be seen that once the window size is reduced to 2,500 the null of local unifractality can no longer be rejected for the EUR/USD and JPY/USD exchange rate return series in some sub-periods. Decreasing the window size further to 1,000 as in panels (b) and (d) further increases the number of windows in which the null of local unifractality cannot be rejected. The results in Figure 6 show that shorter window lengths are also associated with an increase in apparent local unifractality for the

²³For assets that are very liquid, such as those considered here, even higher sampling frequencies, such as 1-minute data, could arguably be used without encountering serious issues with market microstructure effects.

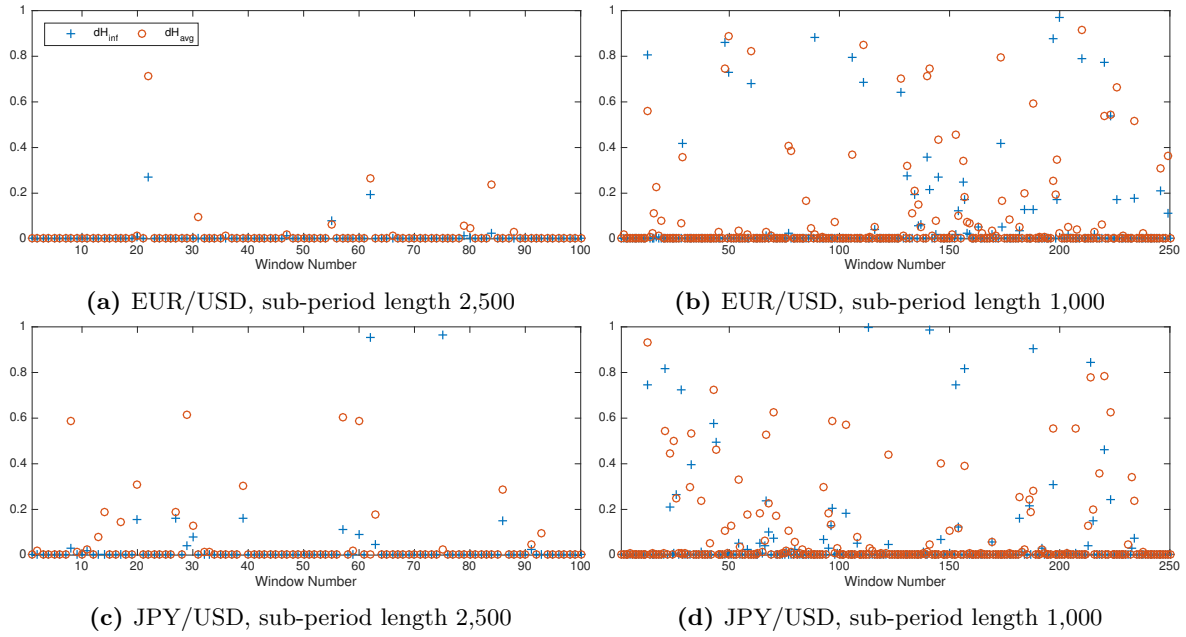


Figure 5: P-values for tests of unifractality versus multifractality applied to consecutive non-overlapping sub-periods of 5-minute return data for the EUR/USD and JPY/USD exchange rates. Crosses and circles indicate sample p-values for the dH_{inf} and dH_{avg} test statistics respectively.

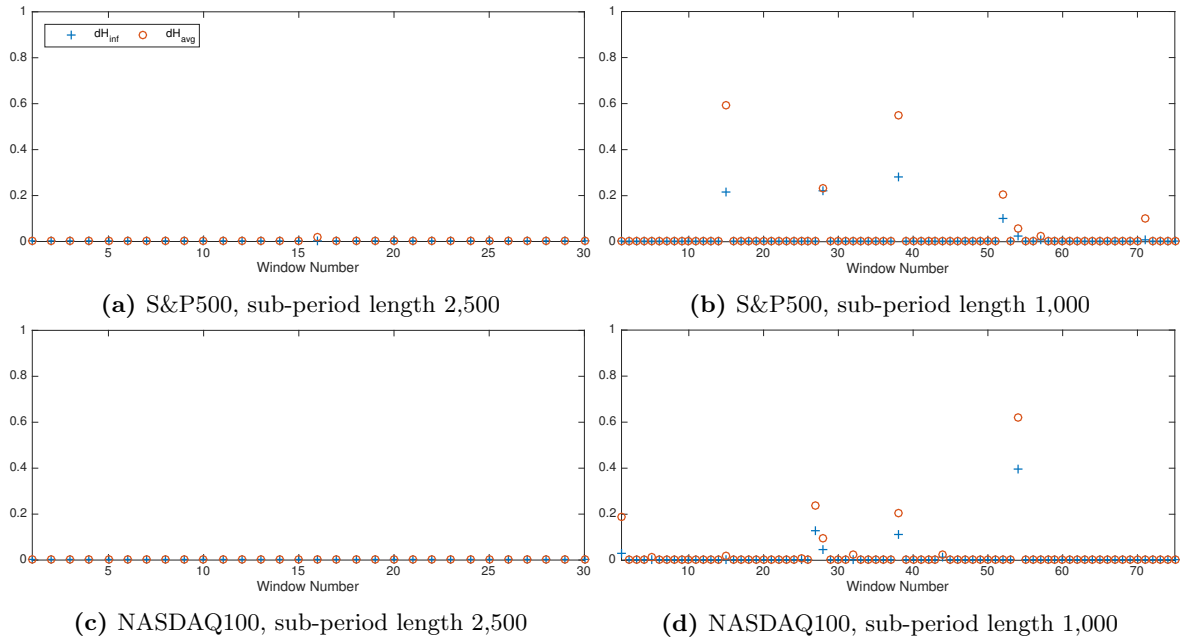


Figure 6: P-values for tests of unifractality versus multifractality applied to consecutive non-overlapping sub-periods of 5-minute return data for the S&P500 and NASDAQ100 equity indexes. Crosses and circles indicate sample p-values for the dH_{inf} and dH_{avg} test statistics respectively.

S&P500 and NASDAQ100 data, however the evidence of local unifractality is much lower, only beginning to appear once the window length is reduced to 1,000.

The first possible explanation for the increasing evidence of local unifractality observed as the window size decreases is that the return processes are in fact truly multifractal in nature, both locally and globally, with the inability to reject the null of unifractality for shorter windows simply due to the power of the tests decreasing as the sample size is reduced. Whilst we cannot rule out this explanation, if it is true then power must depend very strongly on other characteristics of the return processes beyond simply the sample size, since from Figures 5 and 6 the exchange rate series exhibit much stronger evidence of local unifractality than the equity index returns, despite using identical sample sizes.

Following from the earlier discussion of Sections 2 and 3, the second possible explanation for the evidence of local unifractality, when combined with the earlier strong rejection of global unifractality, is that the return series are multifractional rather than multifractal, with this only becoming apparent with shorter window lengths that make the true locally unifractal scaling properties of the return series more visible. One possible justification in support of this explanation is that although windows containing the same number of intraday observations are used for both the exchange rate and equity index series, the large differences in trading hours between these two market types mean that they correspond to very different lengths of trading time; for example, the windows of 1,000 observations represent over twelve trading days of activity for the equity index series, but approximately three trading days for the exchange rate series. If the return series are indeed multifractional rather than multifractal, then the true locally unifractal properties of a given multifractional process should become easier to detect when viewing the process over shorter and shorter time periods. As a result, this possible relationship between the length of the sub-periods in terms of trading time and the evidence of local unifractality could be consistent with the explanation of a multifractional return process, with the variability of the dynamic Hurst exponent for each return series related in some way to trading time.

Whilst intuitively plausible, this hypothesis could in principle be investigated more formally by extending the Monte Carlo exercises of Section 5 to include simulated data from multifractional data generating processes, such as the multifractional Brownian motion. If similar patterns to those in Figures 5 and 6 were observed for simulated multifractional data, then this would add substantial support to this theory. However, due to the practical difficulties with such an exercise previously outlined in Section 5, this is currently left as an avenue for future research

7 Conclusion

The current work has developed a formal statistical testing procedure for determining which subclass of fractal process is most consistent with the distributional scaling properties of a given sample of data. Unlike the only directly equivalent test of Wendt and Abry (2007)

proposed previously in the signal processing literature, our test is developed specifically to be appropriate for realistic financial applications, most notably for testing the distributional scaling properties of financial return series.

When applied to the complete available time series of returns for an asset, our procedure can be employed to distinguish between the type of globally multifractal scaling implied by the multifractal models for asset returns proposed in the financial econometrics literature and the simpler class of unifractal processes. We also discuss how our testing methods can be applied to sub-samples of the complete sample, to test the local distributional scaling properties of the return series and thus potentially distinguish between multifractal processes and the multifractional processes that have begun to attract interest more recently as an alternative approach for modelling return series.

Our proposed testing methodology consists of a set of 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 or critical values for the statistics. The proposed test statistics measure the size of the deviation from unifractality nonparametrically through the use of numerical differentiation methods, thus avoiding the need to specify a particular multifractal process under the alternative hypothesis. In contrast to the earlier informal testing methods employed in the finance literature, the model-based bootstrap resampling scheme takes into account the uncertainty in the sample estimates of the scaling properties, producing p-values that account for the properties of the chosen estimation method.

Monte Carlo exercises using simulated data demonstrate that the proposed testing methodology has good size and power properties in a wide range of relevant testing environments for financial applications and outperforms the competing wavelet-based tests of Wendt and Abry (2007) from the signal processing literature. The power of the tests remains high for the two most plausible processes for financial returns under the alternative of multifractality, even when the strength of multifractality is much weaker than that typically exhibited by financial data. Furthermore, strong performance is maintained even in sample sizes that are small for the literature on fractal processes, supporting the use of the method for tests of local scaling properties and thus for distinguishing between multifractal and multifractional distributional scaling.

Finally, the testing methodology is applied to a dataset of intraday exchange rate and equity index data. 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 global scaling properties of the data are consistent with a multifractal rather than unifractal process. However, when the tests are applied to shorter sub-periods of the intraday return series we find evidence of local unifractality for some sub-periods, with the strength of this apparent local unifractality increasing as the length of these sub-periods is decreased.

References

- P. Abry, V. Pipiras, and H. Wendt. Extreme Values, Heavy Tails and Linearization Effect : a Contribution to Empirical Multifractal Analysis. *GRETSI 2007 Conference Proceedings*, 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.
- 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.
- J. Bardet and I. Kammoun. Asymptotic Properties of the Detrended Fluctuation Analysis of Long-Range-Dependent Processes. *IEEE Transactions on Information Theory*, 54(5): 2041–2052, 2008.
- 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.
- S. Bianchi and A. Pianese. Modelling stock price movements: multifractality or multifractionality? *Quantitative Finance*, 7(3):301–319, 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*, 177(2):320–342, 2013.
- T. Conlon, J. Cotter, and R. Gençay. Commodity futures hedging, risk aversion and the hedging horizon. *The European Journal of Finance*, 22(15):1534–1560, 2016.
- S. Corlay, J. Lebovits, and J. L. Véhel. Multifractal stochastic volatility models. *Mathematical Finance*, 24(2):364–402, 2014.
- 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.
- Y. Fan and R. Gençay. Unit Root Tests with Wavelets. *Econometric Theory*, 26(5):1305–1331, 2010.
- 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. Gençay and D. Signori. Multi-scale tests for serial correlation. *Journal of Econometrics*, 184(1):62–80, 2015.
- L. Giraitis and D. Surgailis. CLT and other limit theorems for functionals of Gaussian processes. *Probability Theory & Related Fields*, 70(2):191–212, 1985.
- 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 G. Pamuła. Multifractal background noise of monofractal signals. *Acta Phys. Pol. A*, 121:B34–B39, 2012.
- P. Hall, W. Härdle, T. Kleinow, and P. Schmidt. Semiparametric bootstrap approach to hypothesis tests and confidence intervals for the Hurst coefficient. *Statistical Inference for Stochastic Processes*, 3(3):263–276, 2000.
- M. Hallam and J. Olmo. Semiparametric Density Forecasts of Daily Financial Returns from Intraday Data. *Journal of Financial Econometrics*, 12(2):408–432, 2014.

- 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.
- 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.
- S. Laurent, C. Lecourt, and F. C. Palm. Testing for jumps in conditionally Gaussian ARMA–GARCH models, a robust approach. *Computational Statistics and Data Analysis*, 100 IS-:383–400, 2016.
- S. S. Lee and P. A. Mykland. Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. *Review of Financial Studies*, 21(6):2535–2563, 2008.
- 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.
- T. Lux and L. Morales-Arias. Forecasting volatility under fractality, regime-switching, long memory and student- innovations. *Computational Statistics and Data Analysis*, 54(11): 2676–2692, 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.
- G. M. Molchan. Scaling exponents and multifractal dimensions for independent random cascades. *Communications in Mathematical Physics*, 179:681–702, 1996.
- 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.
- 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: Statistical Mechanics and its Applications*, 390(14):2637–2654, 2011.

- M. Taqqu. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Probability Theory and Related Fields*, 31(4):287–302, 1975.
- 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 A: Statistical Mechanics and its Applications*, 323:578–590, 2003.
- Y. Xue, R. Gençay, and S. Fagan. Jump detection with wavelets for high-frequency financial time series. *Quantitative Finance*, pages 1–18, 2014.

8 Appendix A: Estimation of Distributional Scaling Properties

There are numerous methods available in the literature for the purpose of estimating $H(q)$ or $\tau(q)$ from the sample of interest (see Kantelhardt, 2009, for a survey of many common estimators). The multifractal detrended fluctuation analysis (MF-DFA) estimator of Kantelhardt et al. (2002) is employed here, primarily because it is arguably the most common choice of estimator for financial applications.

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 (log) price series, or equivalently the log price series may be used directly if the cumulative summation in Step 1 is omitted. The MF-DFA estimate of the generalised Hurst exponent is then obtained as follows:

1. Calculate the cumulative sum or ‘profile’ series $\{y_1, \dots, y_T\}$ from the observations of the original series $\{x_1, \dots, x_T\}$ as $y_t \equiv \sum_{s=1}^t x_s$.
2. Divide the profile series y_t into non-overlapping segments of equal length s . Given that T will not generally be exactly divisible by s , the number of segments obtained is the integer part of T/s , which we denote by $[T/s]$, thus discarding some observations at the end of the series²⁴.
3. The local linear trend²⁵ is calculated for y_t over each of the $[T/s]$ segments of length s , with the value of the fitted trend for the i -th observation in segment $\nu = 1, \dots, [T/s]$ denoted by $\tilde{y}_{\nu,i}$. For each of the $[T/s]$ segments the local linear trend is then used to detrend the profile series and the variance of the detrended series over each of the segments is then calculated as:

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

4. These $[T/s]$ detrended variance terms are then used to obtain the value of the q -th order fluctuation function, $F_{s,q}$, at the chosen value of s according to:

²⁴If desired, to avoid discarding observations steps (1) and (2) can be repeated starting from the end of the original series x_t , producing $2[T/s]$ 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 should typically be avoided.

²⁵It is also possible to use higher order polynomial detrending, allowing more complex forms of nonstationarity to be removed from the series. However, it is standard in the finance literature to employ simple linear detrending due to the computational demands of higher order detrending.

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

for any $q \in \mathbb{R} : q \neq 0$.

5. Steps 2 to 4 are repeated for each of the m segment sizes in the vector $\mathbf{S} = \{s_1, \dots, s_m\}$, where $s_1 < \dots < s_m$ and $m \in \mathbb{N} : m \neq 1$, to obtain a set of points on the q -th order fluctuation function, $F_{s_1,q}, \dots, F_{s_m,q}$.
6. For a unifractal or multifractal process the q -th order fluctuation functions satisfy the relationship $F_{s,q} \propto s^{H(q)}$, or equivalently $\log(F_{s,q}) = a + H(q)\log(s)$, where $H(q)$ is the generalised Hurst exponent. An estimate of $H(q)$ for a specific value of q can therefore be obtained from the estimated slope coefficient of a linear regression of $\log(F_{s_i,q})$ on $\log(s_i)$ using the observations $i = 1, \dots, m$. Finally, an estimate of $\tau(q)$ can be obtained via the previous identity $\tau(q) \equiv qH(q) - 1$.

Whilst there is no formal method for choosing the set of window sizes, \mathbf{S} , previous studies (see for example Bashan et al., 2008) have found using simulated data that the inclusion of either very short windows or windows that are large relative to the sample size generally lead to higher average estimation error for $H(q)$. Approximate limits for the optimal minimum and maximum window sizes, s_1 and s_m , have been suggested that are in the ranges of $5 < s_1 < 15$ and $T/10 < s_m < T/5^{26}$.

The above procedure produces an estimate for the generalised Hurst exponent, $H(q)$, at a single prespecified value of q ; a special case occurs at the value of $q = 2$, for which the multifractal DFA estimator simplifies to the (unifractal) DFA estimator for the scalar simple Hurst exponent, H of a unifractal process²⁷. Clearly an assessment of the concavity of $\tau(q)$ or the dependence of $H(q)$ on q cannot be made based on a single estimated point on these functions, however a set of estimated points on the function $H(q)$ can be obtained by repeating Steps 4 to 6 for multiple values of q , with the set of values for q denoted by the k -dimensional vector $\mathbf{q} = \{q_1, \dots, q_k\}$. In general the only condition that must be satisfied by \mathbf{q} is that $\mathbf{q} \in \mathbb{R}^k$, although for the specific test statistics proposed here the values of the k elements of \mathbf{q} must also be equally spaced and inside a compact set. This process produces sets of estimated points $\{\hat{H}(q_1), \dots, \hat{H}(q_k)\}$ or $\{\hat{\tau}(q_1), \dots, \hat{\tau}(q_k)\}$, from which we can construct test statistics to test whether the properties of the sample are most consistent with a unifractal or multifractal data generating process.

²⁶Values in this range were used as a starting point for the Monte Carlo studies of Section 5 and were found to produce tests with good performance in the current context.

²⁷Because for any unifractal process $H(q) = H$ for any $q \in \mathbb{R}$, an estimate of H can in principle be obtained from the MF-DFA estimator using any $q \in \mathbb{R}$, but the value of $q = 2$ simplifies the expression in Step 4 and so is used in the unifractal case.

Appendix B: Simulation of the Multifractal Random Walk Process in Discrete Time

Following Bacry et al. (2001), a discrete time formulation of the MRW $X_{\Delta t, t}$ with time discretisation step²⁸ Δt 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:

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

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 $E[\omega_{\Delta t, k}] = -\lambda^2 \ln(T/\Delta t)$.

As constructed above, the MRW process is linearly independent, but will exhibit non-linear dependence except in the limiting unifractal case of $\lambda^2 = 0$. The process can however be generalised to allow for linear dependence by replacing the simple Gaussian noise $\epsilon_{\Delta t, k}$ with a fractional Gaussian noise with Hurst exponent $H \neq 0.5$, as described by Bacry et al. (2001).

²⁸The 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.