DEPARTMENT OF ECONOMICS

COLLEGE OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

A New Procedure to Test for *H* Self-Similarity

by William Rea, Chris Price, Les Oxley and Marco Reale

WORKING PAPER

No: 16/2008

Department of Economics
College of Business and Economics
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

A New Procedure to Test for H Self-Similarity

by William Rea[1], Chris Price[2], Les Oxley[3], and Marco Reale[2]

September 12, 2008

Abstract

It is now recognized that long memory and structural change can be confused because the statistical properties of times series of lengths typical of many financial and economic series are similar for both models. We propose a new test aimed at distinguishing between unifractal long memory and structural change. The approach, which utilizes the computationally efficient methods based upon Atheoretical Regression Trees (ART), establishes through simulation the bivariate distribution of the number of breaks reported by ART with the CUSUM range for simulated fractionally integrated series. This bivariate distribution is then used to empirically construct a test. We apply these methods to the realized volatility series of 16 stocks in the Dow Jones Industrial Average. We show the realised volatility series are statistically significantly different from fractionally integrated series with the same estimated d value. We present evidence that these series have structural breaks. For comparison purposes we present the results of tests by Zhang and Ohanissian, Russell, and Tsay for these series.

Keywords: Long-range dependence, strong dependence, global dependence,

JEL Codes: C13, C22

[1] Corresponding Author. Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch 8015, New Zealand. Phone: +64-3-364-2331. Email: wsr12@student.canterbury.ac.nz

[2] Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand

[3] Department of Economics and Finance, University of Canterbury, Christchurch, New Zealand

1 Introduction

It is now widely recognized in the econometric literature that a stylized fact of many financial and economic time series is that they appear to exhibit the property of statistical long memory (Diebold and Inoue, 2001; Granger and Hyung, 2004; Sibbertsen, 2004; Banerjee and Urga, 2005).

The difficulty of distinguishing between long memory and structural change was reviewed by Diebold and Inoue (2001) who suggest '. . . in the sorts of circumstances studied in this paper, "structural change" and "long-memory" are effectively different labels for the same phenomenon . . . '.

Sibbertsen (2004) pointed out that the reason distinguishing between long memory and structural breaks is so difficult is because their finite sample properties are similar and so standard methodologies fail. Structural break detection and location techniques tend to report breaks when only long memory is present. Indeed, Wright (1998) proved that the probability that the standard CUSUM test (Brown et al., 1975) would report a break in a long memory time series converged to one with increasing series length. Conversely, long memory estimators tend to report H estimates which indicate long memory when only structural breaks are present even if the series are Markovian.

Nevertheless, many efforts have been made to establish tests or procedures which can reliably determine the difference between, say, a process with genuinely long memory such as a fractionally integrated series and ones for which the long memory property is merely an artifact of incorrect statistical analysis. In this literature some significant papers are Beran (1992), Beran and Terrin (1996, 1999), Teverovsky and Taqqu (1999), Smith (2005), Zhang (2007), and Ohanissian et al. (2008).

In the literature considerable attention has been directed towards understanding the statistical properties of procedures for detecting and quantifying long memory when only structural change is present. The literature on the problem of understanding the statistical properties of procedures for detecting and locating structural change when, in fact, there is only long memory is somewhat sparse, see Sibbertsen (2004) for a survey.

This paper adds to the literature by presenting a study into the statistical properties of one structural break methodology, Atheoretical Regression Trees (ART) (Rea et al., 2006; Cappelli et al., 2008), when applied to simulated long memory time series and applying the new insights to 16 realized volatility series of stocks in the Dow Jones Industrial Average.

The remainder of this paper is set out as follows. Section (2) gives a brief overview of the competing models. Section (3) presents the methods used. Section (4) presents a representative selection of results. Section (5) presents an application to stock market realized volatilities. Section(6) contains the discussion and Section (7) concludes.

2 Models

A number of models have been proposed to account for the extraordinary persistence of the correlations across time found in long memory series.

There are two common sets of models applied across long-memory series from diverse fields. One set are true long memory models, in particular, the Fractional Gaussian Noises (FGN) and Fractionally Integrated (FI(d)) processes. The other set are models which are non-stationary, but mean reverting. For simplicity the types of non-stationary models studied are ones in which the time series can be broken into a series of "regimes" within which it is a reasonable assumption that the mean is stationary.

2.1 Fractional Gaussian Noises and Fractionally Integrated Series

Fractional Gaussian Noises (FGNs) were introduced into applied statistics by Mandelbrot and van Ness (1968) and are the stationary increments of an Gaussian H-self-similar stochastic process.

Definition 1 A real-valued stochastic process $\{Z(t)\}_{t\in\mathcal{R}}$ is self-similar with index $H > 0$ if, for any $a > 0$,

$$\{Z(at)\}_{t\in\mathcal{R}} =_d \{a^H Z(t)\}_{t\in\mathcal{R}}$$

where $\mathcal{R}$ denotes the real numbers and $=_d$ denotes equality of the finite dimensional distributions. H is also known as the Hurst parameter.

Definition 2 A real-valued process $Z = \{Z(t)\}_{t\in\mathcal{R}}$ has stationary increments if, for all $h \in \mathcal{R}$

$$\{Z(t+h) - Z(h)\}_{t\in\mathcal{R}} =_d \{Z(t) - Z(0)\}_{t\in\mathcal{R}}.$$

It follows from Definition 1 that H is constant for all subseries of an H-self-similar process.

FGNs are a continuous time process. Independently Granger and Joyeux (1980) and Hosking (1981) obtained the discrete time counter-parts to FGNs, the Fractionally Integrated processes (FI(d)).

FI(d)s are also a generalization of the "integration" part of the Box-Jenkins ARIMA (p,d,q) (AutoRegressive Integrated Moving Average) models to non-integer values of the integration parameter, d. Denoting the backshift operator by B, the operator $(1-B)^d$ can be expanded as a Maclaurin series into an infinite order AR representation

$$(1-B)^d X_t = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} X_{t-k} = \epsilon_t \tag{1}$$

where $\Gamma(\cdot)$ is the gamma function, and ϵ_t is a shock term drawn from an $N(0,\sigma^2)$ distribution. The operator in Equation (1) can also be inverted and X_t written in an infinite order MA representation. The two parameters H and d are related by the simple formula $d = H - 1/2$.

ARIMA models with non-integer d are known as AutoRegressive Fractionally Integrated Moving Average (ARFIMA) models. The AR(p) and MA(q) parameters in ARFIMA models may be used to model any additional short-range dependence present in the series. Both FGNs and FI(d)s have been extensively studied. See the volumes by Beran (1994), Embrechts and Maejima (2002), and Palma (2007) and the collections of Doukhan et al. (2003) and Robinson (2003) and the references therein.

2.2 Constrained Non-Stationary Models

Klemes (1974) argued that long memory in hydrological time series was a statistical artifact caused by analyzing non-stationarity time series with statistical tools which assume stationarity. Often series which display the long memory property are constrained for physical reasons to lie in a bounded range. But beyond that we have no reason to believe that they are stationary. For example, in the series we study in this paper (realized volatilities, see Section 5), as long as the companies remain in the index their stock price volatilities cannot have an unbounded increasing or decreasing trend.

Models of this type which have been proposed typically have stochastic shifts in the mean, but overall are mean reverting about some long term average.

We define the break model as follows:

$$\mu_{y_t} = \sum_{i=1}^{p} I_{t_{i-1}<t\leq t_i}\mu_i + \epsilon_t \tag{2}$$

where μ_{y_t} is the mean of the time series, $I_{t\in R}$ is an indicator variable which is 1 only if $t \in R$ and 0 otherwise, t is the time, t_i, $i = 1, \ldots, p$, are the breakpoints and μ_i is the mean of the regime i and ϵ_t is an error term which may have serial correlation. In this case, a regime is defined as the period between breakpoints.

It is important to note that Equation (2) is just a way to represent a sequence of different models (i.e. models subject to structural breaks). This model only deals with breaks in mean. It can be generalized for any kind of break. We are considering this class of model because it has been used by others when studying long memory processes. However, given a true break each regime must be modeled separately.

3 Method

3.1 Univariate Mean Number of Breaks Reported by ART

Zhang (2007) determined that the number of breaks reported by ART when applied to Fractional Gaussian Noises was well-described by a Poisson distribution when the series length and the H parameter were fixed. The method presented here is intended to extend his work to obtain a way of gaining a reasonable estimate of the mean number of reported breaks for a wide range of d values and series lengths in fractionally integrated series.

We simulated FI(d) series using `farimaSim` from the package `fSeries` (Wuertz, 2005) in `R` (R Development Core Team, 2005) with lengths from 1,000 to 16,000 data points in steps of 1,000 data points, and d values between 0.02 and 0.48 in steps of 0.02 d units. ART was applied to each series using functions implemented in the `tree` package (Ripley, 2005) and the number of breaks, their locations and associated regime lengths were recorded. For each set of parameter values 1,000 replications were run. This yielded a grid of 384 data points to which we fitted a function.

3.2 Bivariate CUSUM vs ART

We can obtain a bivariate null distribution for FI(d) series of the CUSUM range (Page, 1954; Brown et al., 1975) and breaks reported by ART under the assumption of true long memory through simulation. We simulated 1000 FI(d) series of desired length and d value and estimated the number of report breaks and the range of the CUSUM test using functions implemented in `strucchange` (Zeileis et al., 2002). In the CUSUM test the residuals are standardized by dividing by the estimated series standard deviation and the cumulative summation of the residuals is plotted against time. Under the null hypothesis of no structural breaks in the mean, the cumulative summation forms a Brownian Bridge usually referred to as the empirical fluctuation process (EFP). The range of the CUSUM test is simply the difference between the maximum and minimum values of the EFP.

We plotted the bivariate distribution of these two structural break methodologies.

In the application in Section (5) each realized volatility series yielded a single data point of number of breaks reported by ART and the CUSUM range. This data point was overplotted on the same bivariate distribution and its statistical significance estimated.

3.3 Modified Ohanissian et al. (2008) Test

Ohanissian et al. (2008) proposed a test for true long memory by estimating the fractional integration parameter d at various levels of temporal aggregation and determining if d was constant as required by the FI(d) model. Their test used the GPH estimator (Geweke and Porter-Hudak, 1983). The empirical properties of 12 popular estimators of the long memory parameter were investigated by Rea et al. (2008). In the application in Section (5) the realized volatility series are 2539 data points long. The results of Rea et al. (2008) showed that in a series of only 2539 data points only two of the popular estimators of the long memory parameter, namely the Whittle and the estimator of Haslett and Raftery (1989), were sufficiently accurate to be used on series of this length. With each level of temporal aggregation the number of data points in the series halves. Thus in the realised volatility series the lengths of the first three aggregated series are 1269, 634, and 317 data points respectively. Thus while the test of Ohanissian et al. (2008) is theoretically sound the wide confidence intervals of the GPH estimator for series of these lengths mean that the test is very unlikely to reject the null hypothesis of an FI(d) series even if it was false. The results of Rea et al. (2008) showed that for these series the Haslett and Raftery (1989) estimator should only be used to one level of aggregation. For this reason we selected the Whittle estimator as implemented in `fSeries` (Wuertz, 2005) to replace the GPH estimator and obtained confidence intervals for this estimator through simulation. The Whittle estimator reported an estimate for the long memory parameter H rather than the more common d used in much of the economic and financial literature but these two parameters are related by the formula $H = d + 1/2$ as stated in Section (2) above. We test, as did Ohanissian et al. (2008), whether the long memory parameter H is constant in these series by checking if the 95 percent confidence intervals overlapped at all levels of aggregation.

4 Results

For reasons of space we report a representative selection of results. The remainder are available on request from the authors.

The distribution of the number of breaks reported by ART for series with 4,000 data points is presented in Figure (1). As Zhang (2007) noted, when the d parameter increased, the simulated series underwent a transition from ART reporting no breaks to reporting multiple breaks for all replications.

The mean number of reported breaks per series for various series lengths and d values is presented in Figure (2).

We fitted a function to the empirical data to obtain formulas for calculating the mean and various tail probabilities. The approximations are calculated by computing the following variables in the order given:

$$\gamma = \left[[5d - x_3]_+^2 - \ell/x_4\right]_+$$

$$q = x_1\gamma + x_2\gamma^2$$

$$\begin{aligned} Q &= q + x_7[q]_+\sqrt{[q]_+} + x_9\gamma(d - x_{10})^2/\ell + (x_5\gamma + x_6\gamma^2)/\ell + x_8 q\ell \\ &\quad + x_{12}q(\ell - 8500)^2 + x_{11}[2000 - \ell]_+[d - 0.32]_+[0.46 - d]_+ \end{aligned}$$

$$F = [Q]_+ \tag{3}$$

$$f = \text{floor}\,(F) \tag{4}$$

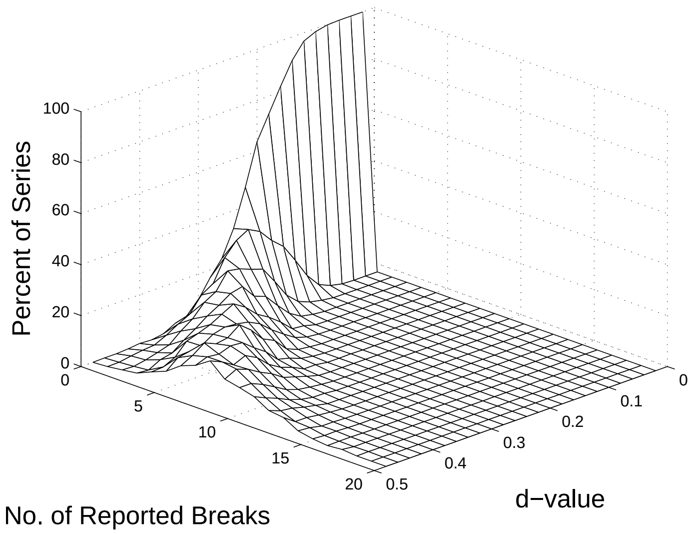

Figure 1: Distribution of the number of breaks reported by ART for 1000 replications of different values of d in an FI(d) series of 4000 data points.

where $x_1 \dots x_{12}$ are given in Table (1). The function $[q]_+$ denotes the maximum of q and zero, and similarly for other arguments. The function 'floor(z)' returns the greatest integer less than or equal to z. When approximating the mean the step in Equation (4) is omitted; $F = [Q]_+$ is used instead.

Each fit has been generated by minimizing a function measuring the error between the fitted function f and the known values $y(d, \ell)$ at data points (d, ℓ), where d ranges from 0.02 to 0.48 in steps of 0.02, and series length ℓ ranges from 1000 to 16000 in steps of 1000. The minimization is with respect to the parameters $x_1, \dots x_{12}$.

The two columns on the left of Table (1) list the parameters for which the probability that the number of breaks is greater than or equal to f is *at least* 97.5% and 95% respectively. In these cases the relevant error function is given by

$$\sum_d \sum_\ell \Big([F(d,\ell) - y(d,\ell) - 1]_+ + [F(d,\ell) - y(d,\ell)]_- \Big)^2 \qquad (5)$$

where $[z]_-$ denotes the minimum of z and zero. This error function imposes no penalty when $y \leq F < y + 1$ because the discrete nature of the floor

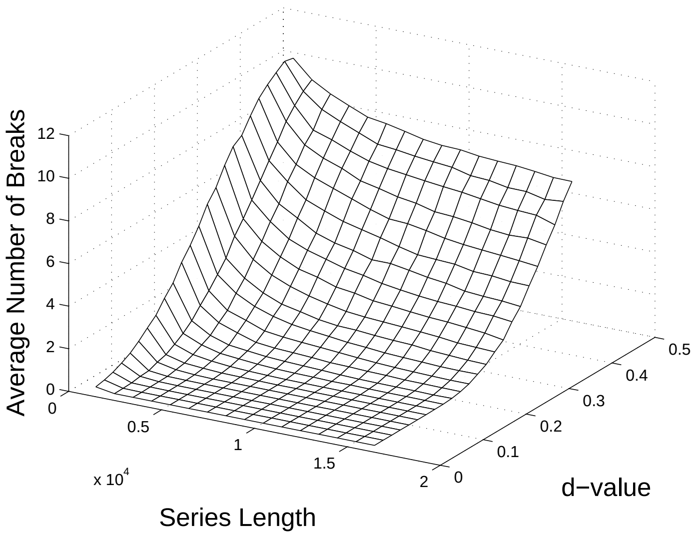

Figure 2: Mean number of breaks reported by ART for different values of d in an FI(d) series with lengths ranging from 1000 to 16000 data points.

function means $f = y$. The use of F rather than f in Equation (5) gives a better measure of the fit in the region between data points. The errors with the optimal parameter values for these fits are 0.0332 and 0.0894 respectively.

The third column of Table (1) lists the parameter values which gives the fitted approximation to the mean. In this case the least squares error is simply

$$\sum_d \sum_\ell (F(d, \ell) - y(d, \ell))^2$$

and the approximation is given by F, not f. The residual sum of squares for the optimal fit was 0.4436.

The four columns on the right of Table (1) list the parameters for which the probability that the number of breaks is less than or equal to f is *at least* 90%, 95%, 97.5% and 99% respectively. For these, the relevant error function is Equation (5). The values of this error function with the optimal parameters are 0.0576, 0.0835, 0.2125, and 0.33 respectively. Clearly the 90% and 95% fits are rather better than the other two. These two fits (90% and 95%) have approximately the same final errors as the two upper quantile fits, however the latter are zero over much larger areas than the 90% and 95%

Table 1: Coefficients for the function approximating the mean, and various upper and lower quantiles.

	2.5%	5%	mean	90%	95%	97.5%	99%
x_1	0.1003	0.8494	-0.6283	-0.3475	-0.2594	-0.3988	-0.6808
x_2	-0.0084	0.416	0.5883	0.3915	0.3876	0.3558	0.3014
x_3	0.8031	1.0362	-0.0613	-0.5026	-0.5384	-0.7014	-0.9815
x_4	10408	17154	24804	23792	23874	21503	19892
x_5	-1505	5468	3188	3195	3350	3233	2672
x_6	415.1	-2677	-503.4	-375.1	-369.8	-326.6	-213.3
x_7	31.179	-0.1075	-0.1248	-0.0977	-0.0975	-0.0969	-0.0937
x_8	1.66e-4	6.8e-5	-4.690e-6	-5.61e-6	-4.06e-6	-4.1e-6	-3.85e-6
x_9	0.2767	-0.5988	-0.6830	-0.6064	-0.6062	-0.3101	-0.3032
x_{10}	0.0544	0.0606	0.0606	0.0606	0.0601	0.0601	0.0550
x_{11}	0.2366	-0.1209	-0.1245	-0.1313	-0.1306	-0.1439	-0.1705
x_{12}	1.9432e-7	6.67e-9	4.6e-10	6.4e-10	2.8e-10	7.6e-10	1.4e-10

fits. Hence the 90% and 95% fits will have smaller relative errors.

The two upper quantiles given by the first two columns allow two sided tests to be performed. When the lower limit on the two sided test is zero, the lower limit effectively says nothing as the number of breaks can not be negative. In such cases a one sided test should be used in place of the two sided test.

Figure (3) presents the differences between the fitted functions estimated upper 95 percent confidence interval and that estimated from the empirical data.

5 Application – Realized Volatilities

The data set comprised the realized volatility of 16 Dow Jones Industrial Average (DJIA) index stocks and were provided by Scharth and Medeiros (2007). The 16 stocks are Alcoa (AA), American International Group (AIG), Boeing (BA), Caterpillar (CAT), General Electric (GE), General Motors (GM), Hewlett Packard (HP), IBM, Intel (INTC), Johnson and Johnson (JNJ), Coca-Cola (KO), Merck (MRK), Microsoft (MSFT), Pfizer (PFE),

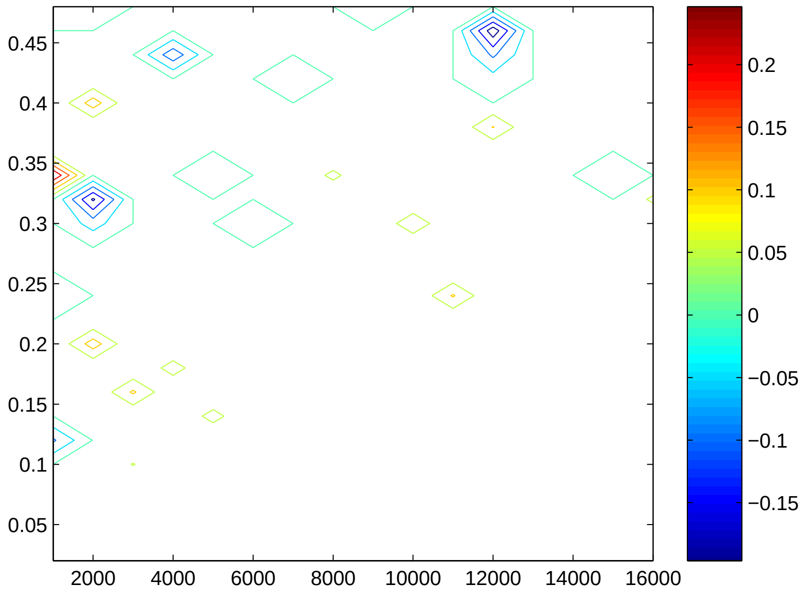

Figure 3: Errors in the fitted function to the empirically determined 95 percent confidence intervals using the formulas in the text. The horizontal axis is the series length, the vertical axis is the d value used in the simulations.

Walmart (WMT), and Exxon (XON). The period of analysis was from January 3, 1994 to December 31, 2003. Trading days with abnormally small trading volumes were excluded, leaving a total of 2539 daily observations. The daily realized volatility was estimated using the two time scale estimator of Zhang et al. (2005) with five-minute grids, which is a consistent estimator of the daily volatility. A fuller explanation of the dataset and how the realized volatilities were calculated can be found in Scharth and Medeiros (2007). It should be noted that because all 16 are part of the DJIA they cannot be considered to be independent series.

We applied the bivariate ART vs CUSUM range as described in the Section (3). For reasons of space we present only a representative selection of results, the remainder are available on request from the authors. The four results are for series with d estimates of 0.36 (GM), 0.40 (JNJ), 0.42 (PFE) and 0.46 (INTC) in Figures (4) and (5).

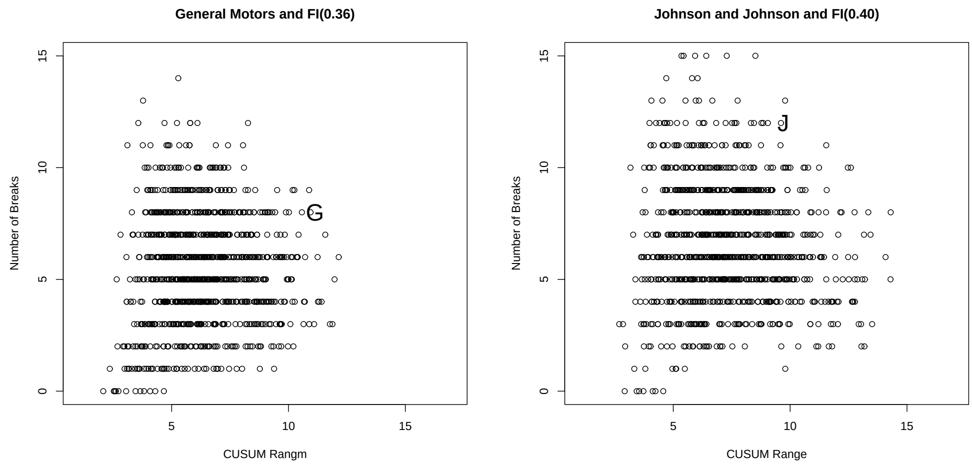

Figure 4: Bivariate distribution of the CUSUM range and number of breaks reported by ART for 1000 replications of different values of d in an FI(d) series of 4000 data points. In the left panel the GM data point is marked with a "G"' and the simulated FI(0.36) data points with open circles. In the right panel the JNJ data point is marked with an "J" the simulated points are FI(0.40).

In Figures (4) and (5) the vertical axis is the number of breaks reported by ART. When considered as a discrete univariate distribution the vertical axis is simply the test of Zhang (2007) (we provide the results of the Zhang test, see Table 2 below). With the exception of JNJ the number of reported breaks in these series was not in the tails of the univariate distribution. Thus the null hypothesis of a fractionally integrated series would not be rejected on the basis of Zhang's univariate test. The horizontal axis is the CUSUM range from the well-known CUSUM test. When taken alone some stocks such as GM and INTC did appear to have a CUSUM range in the tails of this continuous univariate distribution. On a univariate CUSUM test the null hypothesis of a fractionally integrated series would only infrequently be rejected. However, once these two univariate distributions are combined into a bivariate distribution it is clear that the data points from the realized volatilities lie in the tails of the distribution obtained by simulation. For 15 of the 16 realised volatility series the null of a fractionally integrated series was rejected, the exception was WMT.

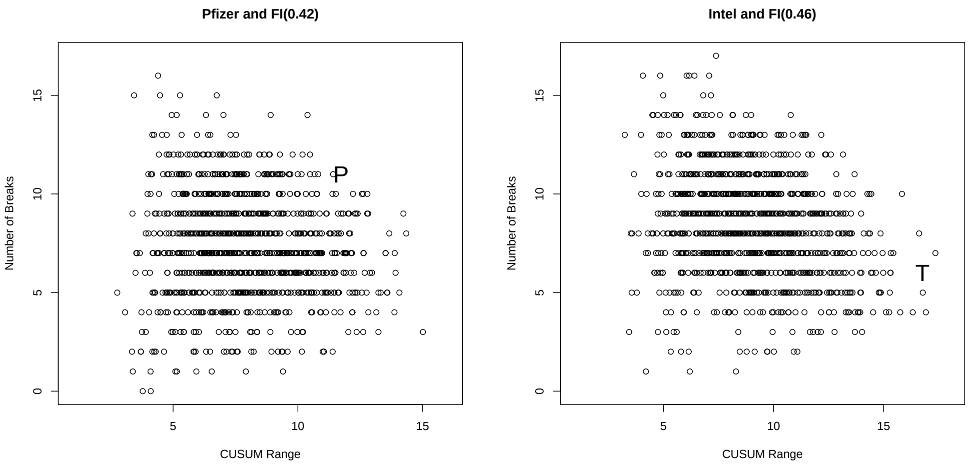

Figure 5: Distribution of the number of breaks reported by ART for 1000 replications of different values of d in an FI(d) series of 4000 data points. In the left panel the PFE data point is marked with an "P"' and the simulated FI(0.42) data points with open circles. In the right panel the INTC data point is marked with an "T" the simulated points are FI(0.46).

For comparison purposes we applied the test of Zhang (2007) and a modified version of the test of Ohanissian et al. (2008).

The results of Zhang's test are presented in Table (2). As can be seen from the Table the null hypothesis of true long memory was only rejected for one of the 16 series at the five percent level, a result which could easily have occured by chance.

Ohanissian et al. (2008) proposed a new test based on comparing the long memory parameter of a series at varying levels of aggregation. In the their test they used the GPH estimator (Geweke and Porter-Hudak, 1983) because of its well understood asymptotic properties which allowed them to theoretically derive critical values for varying levels of statistical significance.

This led us to apply an Ohanissian et al. (2008) type procedure to the 16 stock series using the Whittle estimator rather than the GPH, for reasons outlined in Section (3) above with three levels of aggregation. It should be noted that the series created by three levels of aggregation are only just over 300 data points long. The results are presented in Table (3). With the exception of IBM all the stocks showed an increase in the estimate of the H

Table 2: For the 16 stocks the d estimate is that reported by the estimator of Haslett and Raftery (1989). The actual and expected breaks reported by ART. p-values calculated from Poisson Distribution as in the Zhang (2007) test.

Series	d Est	Reported Breaks	Expected Breaks	p-value
AA	0.42	7	7.5	0.17
AIG	0.40	6	6.8	0.52
BA	0.40	10	6.8	0.08
CAT	0.41	7	7.1	0.52
GE	0.44	8	8.0	0.41
GM	0.36	7	5.5	0.19
HP	0.44	6	8.0	0.69
IBM	0.44	10	8.0	0.18
INTC	0.46	6	8.6	0.75
JNJ	0.40	11	6.8	*0.04
KO	0.42	11	7.5	0.08
MRK	0.39	7	6.4	0.31
MSFT	0.46	10	8.6	0.25
PFE	0.42	10	7.5	0.14
WMT	0.42	10	7.5	0.14
XON	0.44	6	8.0	0.69

Table 3: For the 16 stocks the H estimate is that reported by the Whittle estimator for the full series and three levels of aggregation. The uncertainty is the 95 percent confidence interval. An asterisk (*) indicates a statistically significant change in the H parameter with aggregation at at least the five percent level.

Series	Full	1st Aggregation	2nd Aggregation	3rd Aggregation
AA*	0.858 ± 0.027	0.921 ± 0.042	0.985 ± 0.049	0.994 ± 0.074
AIG*	0.836 ± 0.026	0.895 ± 0.039	0.964 ± 0.049	0.989 ± 0.074
BA*	0.829 ± 0.026	0.881 ± 0.040	0.918 ± 0.057	0.969 ± 0.074
CAT*	0.841 ± 0.026	0.887 ± 0.040	0.952 ± 0.051	0.991 ± 0.074
GE*	0.865 ± 0.027	0.918 ± 0.042	0.961 ± 0.051	0.988 ± 0.074
GM*	0.804 ± 0.024	0.871 ± 0.040	0.912 ± 0.057	0.990 ± 0.074
HP*	0.878 ± 0.028	0.924 ± 0.042	0.977 ± 0.049	0.994 ± 0.074
IBM	0.875 ± 0.028	0.905 ± 0.039	0.969 ± 0.049	0.957 ± 0.081
INTC	0.889 ± 0.028	0.926 ± 0.042	0.955 ± 0.051	0.984 ± 0.074
JNJ*	0.841 ± 0.026	0.879 ± 0.040	0.943 ± 0.051	0.969 ± 0.074
KO*	0.849 ± 0.026	0.915 ± 0.042	0.975 ± 0.049	0.988 ± 0.074
MRK*	0.828 ± 0.026	0.870 ± 0.040	0.939 ± 0.051	0.975 ± 0.074
MSFT	0.878 ± 0.028	0.930 ± 0.042	0.956 ± 0.049	0.959 ± 0.074
PFE*	0.853 ± 0.027	0.893 ± 0.039	0.937 ± 0.051	0.986 ± 0.074
WMT*	0.861 ± 0.027	0.930 ± 0.042	0.980 ± 0.049	0.992 ± 0.074
XON*	0.868 ± 0.027	0.934 ± 0.042	0.978 ± 0.049	0.994 ± 0.074

parameter with higher levels of aggregation leading to a rejection of the null of a fractionally integrated process in 13 of the 16 cases tested.

6 Discussion

As indicated in Section (1) the problem of distinguishing among models with true long memory and other models which display the long memory property is a difficult problem. This paper makes contributions to the literature in three directions:

1. We present a quick method of calculating the expected number of breaks for series of lengths between 1000 and 16000 data points which can then be combined with standard Poisson distribution tables for the Zhang test.

2. We present a bivariate distribution which, in the 16 series examined, appears to easily show the realized volatility series are not FI(d).

3. We present a modified Ohanissian et al. (2008) test which appears to have better power in short series.

The change of behaviour seen in Figure (1) of ART between values of d for which ART reported no breaks and values for which breaks were reported suggests that to distinguish between long memory and regime switching processes at least two approaches are required. Tests or procedures involving ART would only be useful when H or d was sufficiently high that a reasonable number of breaks would be expected to be reported. When H or d was sufficiently low that no breaks would be expected to be reported some alternative method would need to be used. For financial data with a typical d value of about 0.40 and several thousand observations ART should be useful.

Zhang (2007) established that when the series length and d were fixed the number of breaks reported by ART was well described by a Poisson distribution. This lead to two possible tests based simply on the number of reported breaks in a long memory series. We have obtained the mean number of breaks under the null hypothesis of the series being an FI(d) process. If the number of reported breaks in a series under test exceeds the 95% (or other significance) level based on the Poisson distribution then we reject the null of an FI(d) process in favour of a series which had undergone structural change.

Alternatively, we obtained sufficient empirical data through simulation to establish the 95% confidence interval.

The results reported by Zhang (2007) were encouraging but our results with realised volatilities, reported in Table (2), indicated that the problem of the finite sample properties of FI(d) series and series with structural breaks being similar rendered the test of little help in practice.

The results of the modified Ohanissian et al. (2008) test in Table (3) revealed a curious pattern in that the estimate of H rose with increasing level of aggregation for all series with the single exception of a slight drop for IBM between the second and third levels of aggregation. This was despite the fact that the Whittle estimator is known to be downwards biased for short series with high H values. For 13 of the 16 series the change in H estimate was significant at the 95% level between the third level of aggregation and the original series. Thus for these 13 series the null hypothesis of an FI(d) series is not accepted.

The results of looking at the data with a bivariate breaks vs CUSUM range distribution in Figures (4) and (5) is promising and we believe points the way for future progress in this area. With the exception of the Beran (1992) test, tests based on univariate distributions have, in general, not been successful in distinguishing among the proposed models. On these bivariate distributions the real data is clearly in the extreme parts of the tails of the distribution, all four of the results presented here appear to be significant at close to the 0.001 level. Of the 16 series, for 15 of them the null hypothesis of an FI(d) series is not accepted with d as estimated for the full series.

7 Conclusions

Other authors have expressed reservations about the reality of the long memory property exhibited by many financial and economic series. We have proposed a new method based on a bivariate check of the data under which the real data does not fit the distribution obtained for simulated series. The use of bivariate distributions to distinguish between true fractionally integrated series and other series displaying the long memory property appears to be a very promising avenue of future research. In the first application to realized volatilities this methodology did not accept the null hypothesis of true fractionally integrated series for 15 of the 16 series.

There are unresolved statistical issues which merit further research. Firstly, it should be possible to derive theoretically critical values for the Ohanissian et al. (2008) test using the Whittle estimator in place of the GPH. Secondly, in the bivariate approach we have estimated d but then proceeded as if the d value was known *a priori*. Clearly the bivariate distribution is dependent on d and further work needs to be done to establish the usefulness of the approach.

References

Banerjee, A. and Urga, G. (2005). Modelling structural breaks, long memory and stock market volatility: an overview. *Journal of Econometrics*, 129:1–34.

Beran, J. (1992). A Goodness-of-Fit Test for Time Series with Long Range Dependence. *Journal of the Royal Statistical Society B*, 54(3):749–760.

Beran, J. (1994). *Statistics for Long Memory Processes*. Chapman & Hall/CRC Press.

Beran, J. and Terrin, N. (1996). Testing for a change in the long-memory parameter. *Biometrika*, 83(3):627–638.

Beran, J. and Terrin, N. (1999). Testing for a change in the long-memory parameter. *Biometrika*, 86(1):233.

Brown, R. L., Durbin, J., and Evans, J. (1975). Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society Series B*, 37(2):149–192.

Cappelli, C., Penny, R. N., Rea, W. S., and Reale, M. (2008). Detecting multiple mean breaks at unknown points with Atheoretical Regression Trees. *Mathematics and Computers in Simulation*, 78(2-3):351–356.

Diebold, F. X. and Inoue, A. (2001). Long Memory and Regime Switching. *Journal of Econometrics*, 105:131–159.

Doukhan, P., Oppenheim, G., and Taqqu, M. (2003). *Theory and Applications of Long-Range Dependence*. Birkhaüser.

Embrechts, P. and Maejima, M. (2002). *Selfsimilar Processes.* Princeton University Press.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4:221–237.

Granger, C. W. J. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, 11:213–228.

Granger, C. W. J. and Joyeux, R. (1980). An Introduction to Long-range Time Series Models and Fractional Differencing. *Journal of Time Series Analysis*, 1:15–30.

Haslett, J. and Raftery, A. E. (1989). Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource (with Discussion). *Applied Statistics*, 38(1):1–50.

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

Klemes, V. (1974). The Hurst phenomenon - a puzzle? *Water Resources Research*, 10(4):675–688.

Mandelbrot, B. B. and van Ness, J. W. (1968). Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review*, 10(4):422–437.

Ohanissian, A., Russell, J. R., and Tsay, R. S. (2008). True or Spurious Long-Memory? A New Test. *Journal of Business and Economic Statistics*, 26(2):161–175.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.

Palma, W. (2007). *Long-Memory Time Series Theory and Methods.* Wiley-Interscience.

R Development Core Team (2005). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rea, W., Oxley, L., Reale, M., and Brown, J. (2008). The Empirical Properties of Some Popular Estimators of Long Memory Processes. Working Paper 13/2008, Department of Economics and Finance, University of Canterbury.

Rea, W., Reale, M., Cappelli, C., and Brown, J. (2006). Identification of level shifts in stationary processes. In Newton, J., editor, *Proceedings of the 21st International Workshop on Statistical Modeling*, pages 438–441. IDESCAT, Galway.

Ripley, B. (2005). *tree: Classification and regression trees.* R package version 1.0-19.

Robinson, P. M. (2003). *Time Series with Long Memory.* Oxford University Press.

Scharth, M. and Medeiros, M. C. (2007). Asymmetric Effects and Long Memory in the Volitility of Dow Jones Stocks. working paper.

Sibbertsen, P. (2004). Long memory versus structural breaks: An overview. *Statistical Papers*, 45(4):465–515.

Smith, A. D. (2005). Level Shift and the Illusion of Long Memory in Economic Time Series. *Journal of Business and Economic Statistics*, 23(3):321–335.

Teverovsky, V. and Taqqu, M. (1999). Testing for Long-Range Dependence in the Presence of Shifting Means or a Slowly Declining Trend, Using a Variance-Type Estimator. *Journal of Time Series Analysis*, 18(3):279–304.

Wright, J. H. (1998). Testing for a Structural Break at Unknown Date with Long-Memory Disturbances. *Journal of Time Series Analysis*, 19(3):369–376.

Wuertz, D. (2005). *fSeries: Financial Software Collection.* R package version 220.10063.

Zeileis, A., Leisch, F., Hornik, K., and Kleiber, C. (2002). strucchange: An R Package for Testing for Structural Change in Linear Regression Models. *Journal of Statistical Software*, 7(2):1–38.

Zhang, G. Y. (2007). The Distributions of Change Points in Long Memory Processes. In Oxley, L. and Kulasiri, D., editors, *MODSIM 2007 International Congress on Modelling and Simulation*, pages 3037–3043. Modelling and Simulation Society of Australia and New Zealand.

Zhang, L., Mykland, P., and Aït-Sahalia, Y. (2005). A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Date. *Journal of American Statistical Association*, 100:1394–1411.