DETECTION OF SPURIOUS AND REAL BREAKS IN REALIZED VOLATILITY: AN EMPIRICAL STUDY OF THE DJIA

WILLIAM REA, EDUARDO MENDES, LES OXLEY, AND MARCO REALE

Abstract. Granger and Hyung (2004), Diebold and Inoue (2001) and Smith (2005) demonstrate how long memory and structural change can be confused because their finite sample properties are similar. In this paper we present a new approach to detecting multiple breaks in a series. The approach, which utilises the computational efficient methods based upon Atheoretical Regression Trees (ART), also allows categorisation of breaks as either 'spurious' or 'real'. We present empirical examples of the use of the approach utilisin g data on realised volatility from 16 Dow Jones Industrial Average index stock. Particular attention is placed on 5 stocks which exhibit long memory based upon the Beran (1992) test. Some statistical properties of the regimes identified are also considered.

1. Introduction

It is now well known that long memory and structural change are easily confused see [21], [15] and [42]. Distinguishing between long memory and structural change is difficult because their finite sample properties are similar and so standard methodologies often fail, see [41]. It is often the case that structural break detection and location techniques report breaks when only long memory is present; similarly, long memory measurement techniques often report long memory when only structural breaks are present even when the series are Markovian. Theorists may be interested in knowing the statistical properties of procedures for detecting and quantifying long memory when only structural change is present and vice versa, however, recent developments in financial econometrics mean that practitioners are now also acutely interested in these issues.

Engle's paper [18] on ARCH marked the birth of the field of financial econometrics. Bollerslev's generalisation [11] and ensuing computational simplifications opened the literature to a wide audience of applied economists and the area was given extra impetus with the growing availability of high frequency and ultimately 'tick-by-tick' data. However, one of the original problems with the GARCH approach and the Stochastic Volatility (SV) models that followed is that volatility is 'latent' and attempts to measure it via, e.g., daily squared return measures of volatility typically contains significant measurement error. By aggregating 288 squared 5-minute returns, Andersen and Bollerslev [1] demonstrated that volatility becomes essentially measurable, ex-post. Such aggregated high-frequency sums of squared returns have recently been labeled realised volatility. In a series of papers Andersen, Bollerslev, Diebold and Ebens [2] and Andersen, Bollerslev, Diebold and Labys [3] have studied the properties of realised volatility measures as an accurate measure of actual volatility and hence the growth of work on realised volatility in financial and exchange rate markets (see [36] for an excellent survey of the area).

Key words and phrases. Long-range dependence, strong dependence, global dependence, Hurst phenomena. *JEL code.* C22.

An interesting feature that [3], [34], [44], all highlight, is that the typical financial asset has realised volatilities that are fractionally integrated of order d, where d is around 0.4 giving the data a 'long memory' property. Long memory, also known as long-range dependence, strong dependence, global dependence or the Hurst phenomenon after the work of Hurst (1951), has a somewhat longer history than that found in econometrics and popularised by the work of Granger [20]. Hurst [30] considers long memory in hydrological data; Mandelbrot and Wallis [33], likewise in geophysical records, tree-ring indices, earthquake frequencies and sunspot cycles. Examples from economics include stock market volatilities [21] and [22]. Bollerslev and Wright [12] argue that estimates of the degree of fractional integration are unbiased for daily volatility based on intraday returns, whereas they are severely downward biased when estimated from daily squared returns. Diebold and Inoue [15] show that occasional structural breaks can spuriously suggest the presence of long memory. As financial volatility data seems to have occasional, irregular, level shifts, it would seem important to model/detect these shifts when trying to establish the degree of long memory (and vice versa). On the other hand, de Pooter, Martens and van Dijk [35] report that (using S&P 500 realised volatility) level shifts do not account for the long memory characteristics in the data. The fractional integration parameter declines when structural breaks are explicitly modelled, but remains significantly greater than zero.

The aim of this paper is to consider the statistical properties of procedures designed to detect and location of structural change when in fact the series actually exhibit long memory. Long memory processes typically generate observations that may 'look like' structural change, but in fact these breaks are 'spurious'. What we do here is to consider whether we can categorise 'reported' breaks as 'spurious' or 'real' by examining the statistical properties of the 'regimes' reported by the Atheoretical Regression Trees (ART) see Cappelli and Reale (2007). The approach involves using a structural break technique when ART is applied to a pure long memory time series.

The remainder of the paper is organized as follows. Section (2) sets out the competing models and the empirical problem. Section (3) discusses the methods used in this paper. Section (4) describes the data. Section (5) represents the results of investigation and Section (6) gives final concluding remarks. A concise illustration of ART is given in an appendix in section 7.

2. Models and the Empirical Problem

A number of models have been proposed to account for the extraordinary persistence of the correlations across time found in long memory series. Some have enjoyed considerable success in specific fields such as the Granger[20] aggregation model in econometrics.

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 with a non-stationary mean. For simplicity the types of models studied are ones in which the time series can be broken in a series of “regimes” within which it is a reasonable assumption that the mean in stationary. Some examples are structural break and Markov switching models.

2.1. **Fractional Gaussian Noises and Fractionally Integrated Series.**

FGNs (Mandelbrot and van Ness[32]) 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 $=_d$ 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 while Fractionally Integrated series (FI(d)) series(Granger and Joyeux[22], Hosking[29]) are their discrete time counter-parts.

FI(d)s are 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 by B the backshift operator, the integration operator $(1-B)^d$ can be expanded as a Maclaurin series into an infinite order AR representation

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

where $\Gamma(\cdot)$ is the gamma function. The integration operator in Equation (1) can also be written in an infinite order MA representation.

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 short-range dependence present in the series. Both FGNs and FI(d)s have been extensively studied. See the volumes by Beran[8], Doukhan *et al.*[16] and Embrechts and Maejima[17] and the references therein.

2.2. **Structural Breaks in Mean Model.** Klemes[31] argued that statistical long memory in hydrological time series was the result of non-stationarity in the mean. These types of models typically have stochastic shifts in the mean, but overall are mean reverting about some long term average.

We define the break model as follows:

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

where μ_{y_t} is the mean of the time series, $I_{t\in S}$ is an indicator variable which is 1 only if $t \in S$, t is the time, t_i, $i = 1,\ldots,p$, the breakpoint and μ_i is the mean of the **regime** i. In this case, **regime** is defined as the period between breakpoints.

Some comments must be done on this model. First, it is important to note that (2) is just a way to represent a sequence of different models (i.e. models subjected a structural breaks). Second, we are considering only breaks in mean. This model can be generalized for any kind of breaks. In fact, we are considering this class of model just because this is one of the models mainly considered, but given a true break each regime must be modeled separately.

2.3. **Model Confusion.** Despite the obvious theoretical differences between the FGNs or FI(d)s and non-stationary mean models, they are difficult, if not impossible, to tell apart in real data (Granger and Hyung[21], Diebold and Inoue[15]).

Some of this difficulty can be attributed to the fact that FGNs with $0.5 < H < 1$ and FI(d) processes with $0 < d < 0.5$ exhibit some properties of non-stationary series while being stationary. Similarly, while mean shifting models have a non-stationary mean they exhibit some properties of stationary series, particularly in that they typically do not deviate greatly from the series mean. Thus it may help, at an intuitive level, to think of both classes of model as being approximately stationary.

A number of estimators of H utilize the series' estimated spectral density. For example the periodogram, modified periodogram, and Whittle estimators (Taqqu *et al.*[43]) are all based on the series periodogram. The periodogram is a representation of the linear dependence properties of a time series. As regime switching processes are non-linear, the estimates of H and other linear statistical measures, such as the ACF, are meaningless. See [42] for details of the effects of regime switching on the widely-used Geweke Porter-Hudack estimator.

As alluded to above, the use of structural break detection and location methods, which are a prelude to formulating a regime shifting model, are problematic because they tend to find breaks in FGNs and FI(d) series even though the data generating process is uniform throughout the series. For example, when the standard cumulative summation (CUSUM) test(Brown *et al.*[13]) for detecting structural breaks is applied to long memory series the probability of finding a break converges to one with increasing series length (Sibbertsen[40]). Diebold and Inoue[15] show that a series containing only structural breaks can be confused with long memory if the probability of a break decreases with increasing sample size. In real data sets the probability of a break is a property of the data generating process, which is independent of the sample size. Thus if enough data could be gathered in principle it is possible to distinguish between long memory and structural breaks using existing techniques. However there is often insufficient data to make this distinction.

Despite this risk of model misspecification we could find no empirical study of the statistical properties of the "regimes" in FGNs or FI(d) series of finite sample size when they were incorrectly analyzed by applying structural break location methods to them. The use of ART, a computationally very fast structural break method, has allowed large scale simulation studies, such as this one, to proceed which would have been computationally impractical with established techniques such as that due to Bai and Perron[4][5].

3. Method

In this section we present the methodology used to find and decide if the breaks are real or spurious. The methodology can be summarized in the following two steps: (i) apply a fast technique to detect the breaks; and (ii) given each break period, test whether the break is spurious or not. Both steps are discussed in this section.

A fast technique called Atheoretical Regression Trees (ART), recently introduced by Cappelli and Reale[14], is employed to find the breakpoints. An Atheoretical Regression Tree is a non-parametric approach proposed to find breaks in the level of a stochastic process. This method exploits, in the framework of least square regression trees, the contiguity property introduced by Fisher[19] for grouping a single real variable. See appendix for an exposition about ART.

Given each break period, we then test the null hypothesis that the **break is spurious** against the alternative that the **break is a true structural change**. The probability distribution is obtained empirically by simulating series with the same length and fractional integration parameter of the original series and then estimating breaks (using ART), the size of each break period and its fractional parameter.

Formally, assume $y^{(1,T)} = \{y_t\}_{t=1}^{T}$ is a realization of a stochastic process and $B = \{t_1, t_2, \ldots, t_p\}$ the set of breakpoints identified by ART; so, the series is divided into $p+1$ sub-series. Denote any sub-series i as $y^{(t_{i-1},t_i)}$, $i = 0, 1, \ldots, p+1$ with $t_0 = 1$ and $t_{p+1} = T$. Then, define $L = \{l_1, \ldots, l_{p+1}\}$ and $D = \{d_1, \ldots, d_{p+1}\}$ respectively as the sets of length and fractional integration parameters for each sub-series, and d the fractional integration parameter for the whole process. The null and alternative hypothesis as follows:

$$H_0 \ : \ y^{(t_{i-1},t_i)} \text{ is generated by a spurious break}$$
$$H_a \ : \ y^{(t_{i-1},t_i)} \text{ is generated by a true structural break on } y^{(1,T)}$$

To evaluate this hypothesis we test $P[(l_i, d_i) \in I_\alpha] < (1-\alpha)$ where P is a probability measure and I_α is the α-confidence set (equivalently, we check if $(l_i, d_i) \in I_\alpha$). This test is carried out by simulation as described below:

(1) Simulate N true fractional integrated series FI(d) with T observations;
(2) For each series calculate the sets L and D;
(3) Estimate the empirical distribution and the confidence set I_α.
(4) Verify if $(l_i, d_i) \in I_\alpha$.

As the confidence regions usually have irregular contours and the estimator for d exhibits bias for short series, it is preferable to evaluate the hypothesis test graphically (e.g. verify if the point (l_i, d_i) is inside the region defined by I_α).

The real data sets were also subject to the Beran's goodness-of-fit test[1] (see Beran[7]) to evaluate if a fractional integrated process is appropiate. This test was also applied to each sub-series found by ART to evaluate if a model with a time-varying d (e.g. multifractal models) is appropiate.

4. The Data Set

The data set comprise realized volatility and returns of 16 Dow Jones Industrial Average index stock: Alcoa (AA), American International Group (AIG), Boeing (BA), Caterpillar (CAT), General Electric (GE), Hewlett Packard (HP), IBM, Intel (INTC), Johnson and Johnson (JNJ), Coca-Cola (KO), Merck (MRK), Pfizer (PFE), Wal-Mart (WMT), and Exxon (XON). The period of analysis is from January 3, 1994 to December 31, 2003. Trading days with abnormally small trading are excluded, leaving a total of 2539 daily observations. The daily realized volatility is estimated using the two time scale estimator of Zhang *et al.*[45] with five-minute grids, which is a consistent estimator of the daily volatility. A broader explanation about the dataset can be found in Scharth and Medeiros[38].

Tables 1 and 2 show basics statistics for the returns, squared returns, standardized returns, realized volatility, realized standard deviation and log-realized volatility. We present the mean, median, minimum, maximum, standard deviation, skewness and kurtosis for each series.

Some distributional characteristics described by Andersen *et al.*[3] for the realized exchange rate volatility arise in these tables. (i) There is a similarity between the means of the realized variance and the squared returns. It is expected because both are unbiased estimates of the true volatility. However the standard deviation of the realized volatility is much smaller than the standard deviation of the squared returns. (ii) Both realized volatility and realized standard deviation presents excess of kurtosis and a highly asymmetric distribution. On the other hand, the log of realized volatility presents a lower kurtosis and a more symmetric distribution. For this reason, the log of realized volatility is used instead of the realized volatility

[1]The Beran's test was evaluated using functions implemented in the `R`[37] package `longmemo`[39].

Table 1. Descriptive Statistics for Daily DJIA Stocks Return

This table contains the summary statistics for daily return measures. The sample period covers January 3, 1994 to December 31, 2003 (2539 observations). Standardized returns are obtained by dividing the raw returns by the realized standard deviation.

	Mean	Med.	Min.	Max.	Std.dev.	Skew.	Kurt.
Returns							
AA	0.0592	0.0000	-11.7690	13.9790	2.2455	0.2210	5.7682
AIG	0.0621	0.0000	-9.4058	10.2450	1.8781	0.1586	5.6874
BA	0.0309	0.0000	-19.5150	9.4093	2.1330	-0.6411	10.6439
CAT	0.0546	-0.0105	-12.8600	10.5020	2.1109	-0.0589	5.4642
GE	0.0502	0.0000	-11.1580	12.1910	1.8819	0.1137	6.3498
GM	-0.0001	-0.0420	-18.0810	9.7083	2.0923	-0.3201	7.3881
HP	0.0359	0.0000	-20.7230	20.0120	2.8658	0.0072	7.7369
IBM	0.0767	0.0000	-16.8860	12.2970	2.1900	0.1845	7.6796
INTC	0.0812	0.0625	-24.0190	18.2580	3.0129	-0.3959	8.0512
JNJ	0.0606	0.0101	-17.9610	7.8484	1.6230	-0.4967	10.5365
KO	0.0332	0.0000	-10.8050	9.7942	1.7119	-0.0836	6.3963
MRK	0.0404	0.0000	-9.8196	8.7065	1.8106	-0.1102	5.4244
MSFT	0.0845	0.0000	-16.8570	17.2960	2.3863	-0.1068	7.1315
PFE	0.0712	0.0000	-11.3540	7.6380	1.9601	-0.1518	4.8073
WMT	0.0505	0.0000	-11.1420	10.7670	2.0729	0.0935	5.4565
XON	0.0367	0.0000	-8.8868	10.5240	1.5304	0.0930	5.9905
Sq. Returns							
AA	5.0439	1.6259	0.0000	195.4124	11.0386	7.3733	85.7299
AIG	3.5296	1.0449	0.0000	104.9600	7.6555	5.9224	53.4298
BA	4.5487	1.2410	0.0000	380.8352	14.0992	15.2789	340.6905
CAT	4.4570	1.4066	0.0000	165.3796	9.4090	7.0022	83.7160
GE	3.5425	1.0791	0.0000	148.6205	8.2009	7.5364	88.9029
GM	4.3761	1.4354	0.0000	326.9226	11.0627	14.4255	348.2239
HP	8.2111	2.3333	0.0000	429.4427	21.3148	9.2785	134.9611
IBM	4.8002	1.3652	0.0000	285.1370	12.4219	9.2677	142.2279
INTC	9.0807	2.9856	0.0000	576.9124	24.0322	11.5904	207.8825
JNJ	2.6367	0.8942	0.0000	322.5975	8.1032	26.2985	980.0299
KO	2.9305	0.9539	0.0000	116.7480	6.8030	7.8313	95.5194
MRK	3.2785	1.1291	0.0000	96.4245	6.8879	6.6272	65.6423
MSFT	5.6991	1.8597	0.0000	299.1516	14.0855	11.0667	189.1471
PFE	3.8454	1.3809	0.0000	128.9133	7.4782	6.1214	66.7017
WMT	4.2979	1.3600	0.0000	124.1442	9.0812	5.7336	50.3953
XON	2.3424	0.7988	0.0000	110.7546	5.2368	9.1377	139.4680
Std. Returns							
AA	0.0220	0.0000	-6.3106	3.7788	1.0193	0.0468	3.4540
AIG	0.0536	0.0000	-3.6679	3.7159	0.9875	0.1749	3.0764
BA	0.0243	0.0000	-16.9699	3.0882	1.0041	-1.8581	34.7971
CAT	0.0330	-0.0063	-10.1037	4.0079	1.0468	-0.2529	6.3467
GE	0.0720	0.0000	-3.4612	3.9464	1.0196	0.2564	2.9943
GM	-0.0330	-0.0268	-119.8759	3.7493	2.5941	-38.8369	1795.4531
HP	0.0596	0.0000	-5.2363	7.8799	1.1320	0.3334	5.4140
IBM	0.0626	0.0000	-4.3916	5.6842	1.1232	0.2628	3.5643
INTC	0.1015	0.0348	-9.5360	4.9895	1.3823	-0.0861	4.3746
JNJ	0.0508	0.0104	-3.1235	3.5500	0.9239	0.1542	3.1103
KO	0.0386	0.0000	-2.9779	3.5548	0.9016	0.1853	3.1710
MRK	0.0454	0.0000	-4.3590	4.3478	0.9906	0.0900	3.4009
MSFT	0.0890	0.0000	-6.6010	5.9613	1.3170	0.1525	3.4358
PFE	0.0683	0.0000	-3.1276	3.4387	0.9902	0.1745	2.9191
WMT	0.0430	0.0000	-3.3396	3.3880	0.8944	0.1942	3.2411
XON	0.0445	0.0345	-3.2432	3.1444	0.9226	0.0932	2.9438

Table 2. Descriptive Statistics for Daily DJIA Stocks Realized Volatility

This table contains the summary statistics for daily realized volatility measure. The sample period covers January 3, 1994 to December 31, 2003 (2539 observations).

	Mean	Med.	Min.	Max.	Std.dev.	Skew.	Kurt.
Realized Var.							
AA	5.1834	3.4186	-0.5307	50.2230	5.1477	2.5744	11.8982
AIG	3.3285	2.3245	-0.7663	54.4010	3.2353	4.9796	51.5327
BA	4.5087	3.2051	-0.9989	76.9930	4.5544	4.6060	44.1159
CAT	4.4668	3.0989	-0.6955	47.6930	4.3718	3.3203	20.6553
GE	3.3691	2.3398	-0.7288	50.3660	3.5433	4.8747	42.5651
GM	3.8228	2.7456	-0.9773	39.2060	3.5552	3.6436	22.5032
HP	7.2864	4.5536	-0.6106	102.4400	8.0333	3.6999	27.4805
IBM	3.5255	2.4148	-0.6940	54.6810	3.5263	4.2417	35.3268
INTC	5.3452	3.4125	-0.7937	76.4150	5.7693	3.5469	25.0793
JNJ	3.1808	2.3545	-0.7909	51.6580	3.3208	5.4480	48.3564
KO	3.5087	2.5262	-0.6523	50.7080	3.1923	3.9729	34.8622
MRK	3.4671	2.5647	-0.7290	42.7880	3.4093	4.7862	37.7570
MSFT	3.5001	2.4537	-0.8873	44.3320	3.5268	3.9132	30.2009
PFE	4.1549	2.9970	-0.4824	48.5730	4.2476	4.6315	35.0192
WMT	5.2769	4.0421	-0.5867	84.7110	4.9462	5.7714	64.9237
XON	2.7857	1.9414	-0.7253	42.0930	2.8622	4.5468	38.6561
Realized std. dev.							
AA	2.0819	1.8489	0.6851	7.0868	0.9216	1.3650	4.9927
AIG	1.6975	1.5246	0.4834	7.3757	0.6686	1.8001	9.3120
BA	1.9593	1.7903	0.0337	8.7746	0.8186	1.6648	8.3425
CAT	1.9479	1.7604	0.5518	6.9060	0.8202	1.5066	6.3622
GE	1.6934	1.5296	0.5208	7.0969	0.7082	1.9163	9.7138
GM	1.8261	1.6570	0.1508	6.2615	0.6989	1.7955	7.9198
HP	2.4342	2.1339	0.6240	10.1213	1.1669	1.4548	6.2647
IBM	1.7390	1.5540	0.5532	7.3947	0.7083	1.8219	8.4124
INTC	2.0958	1.8473	0.4542	8.7416	0.9763	1.5033	6.2984
JNJ	1.6627	1.5344	0.4573	7.1873	0.6453	2.4188	13.1318
KO	1.7483	1.5894	0.5897	7.1210	0.6725	1.5795	7.3642
MRK	1.7405	1.6015	0.5206	6.5413	0.6618	2.1733	11.0744
MSFT	1.7181	1.5664	0.3357	6.6582	0.7406	1.4882	7.0161
PFE	1.8909	1.7312	0.7194	6.9694	0.7614	2.0085	10.0670
WMT	2.1568	2.0105	0.6429	9.2039	0.7909	2.0199	11.7544
XON	1.5412	1.3933	0.5241	6.4879	0.6407	1.8261	8.9054
Log realized var.							
AA	1.2961	1.2292	-0.7564	3.9165	0.8094	0.3432	2.7195
AIG	0.9274	0.8435	-1.4537	3.9964	0.7069	0.3930	3.2375
BA	1.1917	1.1647	-6.7789	4.3437	0.7828	-0.2094	7.1515
CAT	1.1798	1.1310	-1.1891	3.8648	0.7703	0.2922	3.0029
GE	0.9085	0.8501	-1.3048	3.9193	0.7422	0.4076	3.2374
GM	1.0828	1.0100	-3.7832	3.6688	0.6787	0.4173	4.2865
HP	1.5763	1.5159	-0.9432	4.6293	0.8908	0.1943	2.6934
IBM	0.9699	0.8816	-1.1841	4.0015	0.7174	0.5076	3.2848
INTC	1.2916	1.2274	-1.5784	4.3362	0.8537	0.2634	2.8204
JNJ	0.8994	0.8563	-1.5650	3.9446	0.6606	0.5971	4.2200
KO	0.9901	0.9267	-1.0563	3.9261	0.6972	0.4029	3.0394
MRK	0.9927	0.9418	-1.3056	3.7563	0.6571	0.5757	3.9329
MSFT	0.9157	0.8976	-2.1833	3.7917	0.8133	0.0500	3.1672
PFE	1.1414	1.0976	-0.6586	3.8831	0.7075	0.4698	3.4401
WMT	1.4238	1.3968	-0.8835	4.4392	0.6604	0.2934	3.8115
XON	0.7209	0.6634	-1.2920	3.7399	0.7403	0.4203	3.1252

itself. (iii) the returns are asymmetric and heavy tailed, but the standardized returns are much less asymmetric and leptokurtic, presenting a distribution close to the Normal distribution.

Table 3 presents the estimates of the fractional integrated parameter d using the Whittle estimator and Beran's goodness-of-fit tests for each series. As reported by other authorities a d estimate close to 0.4 is appropriate for most series. Despite the fact that realized volatilities usually present a long memory behavior and a visual examination of the series ACF and periodogram suggests the series are of the long memory type, most of the series (AA, BA, CAT, GE, HP, IBM, INTC, KO, MSFT, PFE and WMT) rejected the Beran's test at a level of 5%. The rejection sugests that a unique d cannot explain the whole series or the series does not have long-memory. However, this analysis is beyond the scope of this paper and we will focus on the series with long memory behavior.

Table 3. Fractional Integrated Parameter estimates for Daily DJIA Stocks Realized Volatility

This table contains the estimates of the fractional integrated parameter d using the Whittle estimator and Beran's goodness-of-fit tests for the whole series.

	d	**Beran p-value.**
AA	0.42	2.65e-6
AIG	0.40	0.14
BA	0.40	0.02
CAT	0.41	0.002
GE	0.44	0.04
GM	0.36	0.09
HP	0.44	0.002
IBM	0.44	8.93e-6
INTC	0.46	3.59e-7
JNJ	0.40	0.16
KO	0.42	2.57e-6
MRK	0.39	0.34
MSFT	0.46	0.008
PFE	0.42	4.43e-6
WMT	0.42	2.95e-12
XON	0.44	0.09

5. Results

We apply the procedure described in section 3 to the five stocks which seem to exhibit long memory (AIG, GM, JNJ, MRK and XON). For all series we evaluate graphically the hypothesis that the breaks (detected by ART) are spurious at $\alpha =$ 95% confidence level. We found that JNJ. MRK and XON are really long memory processes and AIG and JNJ have breaks.

Table 4 presents the test results for each stock at a $\alpha = 95\%$ confidence level. Clearly only two series have breaks and the others are regarded as true long memory process. The breaks found in the AIG and the GM series are dated 20 July, 1998 and 03 December, 1999 respectively. The break found in AIG series can be associated with the Russian Crisis. The break found in the GM series can be associated with several events, for example the selling of GM's Financial Transaction Business to

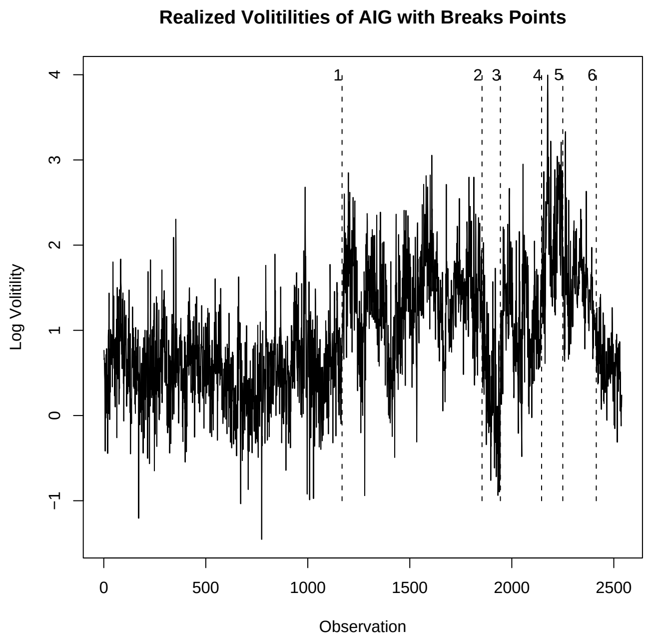

FIGURE 1. Series plot with breaks identified by ART.

Arthur Andersen (23-Nov-99), a cross-supply agreement with Honda (21-Dec-99), and Management Changes (08-Feb-2000). This last event could be regarded as a main event which lead to the break.

TABLE 4. NUMBER OF TRUE BREAKS DETECTED

This table shows the number of detected Breaks and divide them into Spurious Breaks and True Breaks.

	# Breaks	# Spurious Breaks	# True Breaks
AIG	7	6	1
GM	8	7	1
JNJ	11	11	0
MRK	8	8	0
XON	7	7	0

To illustrate the technique we will present an analysis of AIG. First the breaks detected by ART are presented with a possible explanation for each, and then we show the results of the test for AIG.

Figure 1 shows the realized volatility series with each break identified. We can verify that after break 1 the level of the series changes and the series presents a more "irregular" behavior. The period between breaks 1-2 and 3-4 apparently have a behavior similar to the first regime but with a higher level. Moreover, periods between breaks 2-3 and 4-5, and 6- , are likely to be less and more volatile periods respectively.

Some of these breaks can be associated with some events:

(1) 20-July-1998: Russian Financial Crisis.

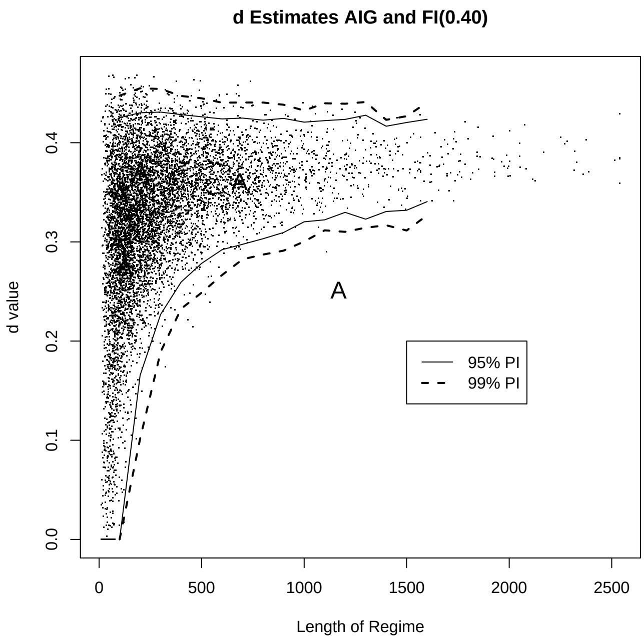

FIGURE 2. Density plot with the 95% confidence set (dashed line). The regimes are represented with the symbol "A". A point outside the confidence set means a rejection of the null hypothesis

(2) 09-April-2001: Talking between AIG and American General Corp. (AGC). In May, 2001, AIG bought AGC for US$ 23 billion.
(3) 16-August-2001: Reasonably close to 09/11 Terrorist ACT.
(4) 12-June-2002: Close to DJIA 4-year Low.
(5) 06-November-2002: No event related.
(6) 03-July-2003: No event related.

According to the test, it appears that only the Russian Crisis caused a structural change in the series. Figure 2 shows the result of the test. This figure is a density plot (a 2-dimension histogram) where the dashed line is the 95% confidence set. We plot the "A" on the same graph and a point outside the dashed region denotes a rejection of the null hypothesis.

6. CONCLUSIONS

Many previous studies have demonstrated the value of FGNs and FI(d) processes as an operational model for stock market log volitility time series. Indeed, if the goal of the analysis is to find a single statistically parsimonious and mathematically elegant model for all of the data then either an FGN or an FI(d) are good choices. They are quick to fit, require few parameters, and are straight forward to use in forecasting. For pragmatic time series analysis FGNs and FI(d) processes provide useful modeling tools.

However, a pure FI(d) process does not fit the observed distributions of a range of statistical properties of the realized log volitility for the AIG stock. This suggests that a more complex physical model of log volitility is required than is offered by an FI(d) process.

7. Appendix: Atheoretical Regression Trees

Time series data consists of a series of observations ordered by time. In the context of regression trees (RTs) time assumes the role of the predictor variable when, in fact, it is merely a counter. A common source of poor predictive performance in RTs is that the distribution of the response variable is not orthogonal to the predictor variables (see Fig 8.12 of [27]) for an example). This problem does not arise in univariate time series. This gives us reason to suspect they will perform well in the location of structural breaks.

There are several questions to be addressed in applying RTs to time series. These are:-

(1) As RTs fit piece wise constant functions to data do RTs discover or impose breaks on time series?
(2) What is the effect of serial correlations on RTs performance in detecting structural breaks?
(3) Given that observations in time series are, in general, non-interchangeable can cross-validation be used in tree selection?
(4) Is it possible to obtain a confidence interval for the breaks?

The model considered is:

$$y_t = \mu_g + \epsilon_t, \quad g = 1, \ldots, G, \quad t = T_{g-1} + 1, \ldots, T_g, \tag{3}$$

where G is the number of regimes (and $G-1$ the number of breakdates), y_t is the observed response variable and ϵ_t is the error term at time t (we adopt the common convention that $T_0 = 0$ and $T_G = T$ where T is the series length). This is a pure structural breaks model because all the model coefficients are subject to change and it has been employed by Bai & Perron [5] to detect abrupt structural changes in the mean occurring at unknown dates. The problem is to estimate the set of breakdates $(T_1, \ldots, T_g, \ldots, T_{G-1})$ that define a partition of the series

$$P(G) = \{(1, \ldots, T_1), \ldots, (T_{g-1} + 1, \ldots, T_g), \ldots, (T_{G-1} + 1, \ldots, T)\},$$

into homogeneous intervals such that $\mu_g \neq \mu_{g+1}$. Bai & Perron (2003) propose an estimation method based on the least squares principle: for each G-partition, the corresponding least square estimates of the μ_g's are obtained by minimizing the within-group sum of squares

$$WSS_{y|P(G)} = \sum_{g=1}^{G} \sum_{t=T_{g-1}+1}^{T_g} (y_t - \mu_g)^2. \tag{4}$$

The estimated breakdates $(\hat{T}_1, \ldots, \hat{T}_g, \ldots, \hat{T}_{G-1})$ are associated with the partition $P^*(G)$ such that $P^*(G) = argmin_{P(G)} WSS_{y|P(G)}$. In this approach, the breakdate estimators are global minimizers since the procedure considers all possible partitions by using the dynamic programming approach proposed by Fisher's (1958) to find the least squares partition of T *contiguous* objects into G groups. His efficient algorithm exploits the additivity of the sum of squares criterion resorting to a dynamic programming approach [6] that applied to ordered data points finds the global minimum. Despite the computational saving, the method cannot deal with high values of T and G and the same remark holds for the Bai & Perron's procedure, even with today's computing power.

In the case of time series data [26] provides an excellent justification in favor of the (faster) binary division algorithm: suppose that the observed time series consists of G segments within each of which the values are constant, i.e. model (1) becomes a piecewise constant model with $\epsilon_t = 0$. Then, there is a partition into G segments for which the within-group sum of squares is zero and it will be identified

by a sequential splitting algorithm as the one in ART.
In other words, if the data have a hierarchical structure then ART will find the overall optimum, otherwise it provides a suboptimal solution for which, because the partitions are contiguous, misplacements can occur only on the boundaries. As discussed in [25], although structural breaks are treated as immediate, it is more reasonable to think that they take a period of time to become effective, thus misplacements on the boundaries are not a concern.
Given that the global search algorithm requires $O(n^2)$ steps, whereas ART, at any tree node requires $O(n(h))$ steps to identify the best split, suboptimality does not appear a high price to pay to obtain full feasibility and indeed, in the application we will show that the partitions provided by ART are comparable to those obtained by the global search procedure.

References

[1] T. Andersen and T. Bollerslev. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39:885–905, 1998.

[2] T.G. Andersen, T. Bollerslev, F.X. Diebold and H. Ebens. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61:43–76, 2001.

[3] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of American Statistical Association*, 96:42–55, 2001.

[4] J. Bai and P. Perron. Estimating and test ng linear models with multiple structural changes. *Econometrica*,66:47–78, 1998.

[5] J. Bai and P. Perron. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18:1–22, 2003.

[6] R.E. Bellmann and S.E. Dreyfus. *Applied Dynamic programming*. Princeton University Press.

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

[8] J. Beran. *Statistics for Long Memory Processes*. Chapman & Hall/CRC Press, 1994.

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

[10] J. Beran and N. Terrin. Testing for a change in the long-memory parameter. *Biometrika*, 86:233, 1999.

[11] T. Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[12] T. Bollerslev and J. H. Wright. Semoparametric estimation of long memory volatility dependencies: the role of high frequency data. *Journal of Econometrics*, 98:81–106, 2000.

[13] R.L. Brown, J. Durbin, and J.M. Evans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society B*, 37(2):149–163, 1974.

[14] Carmela Cappelli and Marco Reale. Detecting Changes in Mean Levels with Atheoretical Regression Trees. Research Report UCMSD 2005/2, Department of Mathematics and Statistics, University of Canterbury, 2005.

[15] F.X. Diebold and A. Inoue. Long memory and regime switching. *Journal of Econometrics*, 105:131–159, 2001.

[16] P. Doukhan, G. Oppenheim, and M. Taqqu. *Theory and Applications of Long-Range Dependence*. Birkhaüser, 2003.

[17] P. Embrechts and M. Maejima. *Selfsimilar Processes*. Princeton University Press, 2002.

[18] R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987-1006, 1982.

[19] W.D. Fisher. On grouping for maximum homogeneity. *Journal of the American Statistical Association*, 53(284):789–798, 1958.

[20] C.W.J. Granger. Long Memory Relationships and the Aggregation of Dynamic Models. *Journal of Economics*, 14:227–238, 1980.

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

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

[23] C.W.J. Granger and T. Teräsvirta. A simple nonlinear time series model with misleading linear properties. *Economics Letters*, 62:161–165, 1999.

[24] J.D. Hamilton. A new approach to the economic analysis of non-stationary time series and the business cycle. *Econometrica*, 57:422–437, 1989.

[25] B.E. Hansen. The new econometrics of structural change: dating breaks in the U.S. labor productivity. *Journal of Economic Perspectives*, 15:117–128, 2001.

[26] J.A. Hartigan. *Clustering Algorithms.* John Wiley & Sons.

[27] T. Hastie, R. Tibshirani and J.H. Friedman. *The Elements of Statistical Learning.* Springer, 2003.

[28] T. Higuchi. Approach to an irregular time series on the basis of fractal theory. *Physica D*, 31:277–283, 1988.

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

[30] H.E. Hurst. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770-779, 1951.

[31] V. Klemes. The Hurst phenomenon - a puzzle? *Water Resources Research*, 10:675–688, 1974.

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

[33] B.B. Mandelbrot and J.R. Wallis. Some long-run properties of geophysical records. *Water Resources Research*, 5:321-340, 1969.

[34] M. Martens. Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. *Journal of Futures Markets*, 22:497-518, 2002.

[35] M. de Pooter, M. Martens and D. van Dijk. Predicting the daily covariance matrix of S&P100 stocks using intraday data - but which frequency to use? *Econometric Reviews*, forthcoming.

[36] M. McAleer and M.C. Medeiros. Realized volatility: a review. *Econometric Reviews*, forthcoming.

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

[38] M. Scharth and M.C. Medeiros. Asymmetric effects and long memory in the volatility of djia stocks. Discussion Paper 532, Pontifical Catholic University of Rio de Janeiro, 2007.

[39] S scripts originally by Jan Beran Datasets via Brandon Whitcher ¡brandon@stat.washington.edu. Toplevel R functions and much more by Martin Maechler. *longmemo: Statistics for Long-Memory Processes (Jan Beran) – Data and Functions*, 2006. R package version 0.9-3.

[40] P. Sibbertsen. Robust CUSUM-M-Test in the Presence of Long Memory Disturbances. Technical Report 19/2000, Universität Dortmund, 2000.

[41] P. Sibbertsen. Long memory versus structural breaks: an overview. *Statistical Papers*, 45:465–515, 2004.

[42] A. D. Smith. Level shift and the illusion of long memory in economic time series. *Journal of Business and Economic Statistics*, 23:321–335, 2005.

[43] M. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: an empirical study, 1995.

[44] D.D. Thomakos and T. Wang. Realized volatility in the futures markets. *Journal of Empirical Finance*, 10:321353, 2003.

[45] L. Zhang, P. Mykland, and Y. Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency date. *Journal of American Statistical Association*, 100:1394–1411, 2005.

(W.Rea and M. Reale) Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

(E. Mendes and L. Oxley) Department of Economics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand