[DRAFT]

Long-term dependency of market returns and MMAR (Multifractal Model of Asset Returns)

Outline:

Abstract:

Section 1: Introduction

Section 2: MMAR

* mathematical construction of MMAR
* its use to satisfy the model

Section 3: Hurst coefficients and R/S method

* What hurst exponent is
* Different applications in other fields
* Different estimator methods
* Difficulty in estimating
* Classic R/S method
* Why we use R/S method (problems with R/S, briefly introduce Lo’s modification)

Section 4: Hurst coefficients for Kenneth French’s portfolios

* 48 industry portfolios with 48 different Hurst coefficients

Section 5: Test hypothesis that there is relationship between Hurst and average return

* 48 coefficients with 48 returns
* correlations?

Section 6: Related studies

Section 7: Conclusion

Section 8: References

Section 1: introduction

The Bachelier-Black-Scholes random walk model has long been documented for its inadequacies in dealing with the rich statistical textures exhibited by the financial time series. Two major problems that exist with this model have been extensively studied. One of these two problems is the well-known fat tail of the logarithmic returns in the empirical data. As noted by countless empirical evidences, the return distribution is in fact strongly non-Gaussian and has a strong kurtosis. The tails generally decay in a power law with tail exponent up to 3 to 5 in liquid markets and even 2 in emerging markets [B.B.Mandelbrot, The variation of certain speculative prices, J. Business 36, 394(1963), also reference of 8,9,5,10 in Bouchaud’s paper]. The other problem exists with its independent white noise assumption. In the original Bachelier model assumption, returns are uncorrelated and even independent and identically distributed. Various problems with this assumption have been discovered, too. For instance, volatility signal in a given time period has found to be correlated and high volatility periods tend to persist in time. This volatility clustering effect has been found in daily and high frequency intraday data. The leverage effect, on the other hand, has also been recently discovered and it shows the market’s past price has a negative correlation with future volatility. [J.P. Bouchaud, A. Matacz, M.Potters, The leverage effect in financial markets: retarded volatility and market panic. Physical Review Letters, 87, 228701 (2001).] Problems like these create difficulties for people to manage financial risk when continue using the classic random walk model. Alternative theories have hence been developed to study interesting non-linear effects with non-Gaussian random walks.

Of the alternatives proposed, the MMAR model introduced by Mandelbrot, Fisher and Calvet in 1997 has received support of more and more empirical evidences from various studies. Ideally one needs a model of asset price return with the following properties: ability to accurately capture the observed leptokurtosis (or the fat tail), ability to incorporate the non-independence of returns over disjoint time periods (or so called volatility clustering), and not a priori excluding the existence of the second moment of the return process[Multifractality in SA rand exchange rates, PH Potgieter, contributed talk at SUMMER SCHOOL IN MATHEMATICAL FINANCE, Inter-University Centre Dubrovnik, Croatia, September 2001.]. The second requirement motivated the development of ARCH and GARCH models, which was studied to be found rather weak to satisfy the first requirement. (The first requirement was satisfied by power law distributions and its existence in real world has been recognized as a stylized fact. A universal inverse cubic power law was postulated by the Boston Group and so far the empirical result has been convincing in the high frequency data when studying price impact functions. \*Paper and reference to be attached). The last requirement is directly related to Mandelbrot’s initial fractal model (1966) where he used levy stable distribution and one of the assumptions is that process could be stationary but with infinite variance. (Professor Nassim Taleb here at Poly still holds this kind of opinion that standard deviation is a meaningless measure facing scale-free extreme events.) However, a finite and converging second moment is still highly desirable for practical purposes. MMAR model that will be studied in this paper satisfies all three requirements.

In section 2 we will briefly review the mathematical construction of MMAR with concepts in mulfitractal measures and fractional Brownian motion. In section 3 we will focus on the Hurst Coefficient H, introduced by Mandelbrot for remembrance of Harrold Hurst, a British hydrologist who discovered fractality in water resource engineering. The Hurst coefficient can be found through MMAR model and measures the long-term dependency of time series. Hurst exponent has been widely applied in various scientific fields characterizing long-term dependency and multifractality. We will also introduce the standard R/S analysis to measure H and compare with other methods in section 3. Then in section 4 we compute the H coefficients for various portfolios and try to interpret its meanings. In section 5 we test the hypothesis that higher H results in higher returns for different portfolios, which has been tested on hedge fund manager selections to be true. Related studies and further reading discussions will be also included in a separate section 6 due to its importance. The paper will conclude in section 7.

Section 2: MMAR (Multifractal model of asset returns)

We will now introduce the construction of MMAR, which originates from Mandelbrot’s fractal theory back in 1966. The mathematics derived from fractal geometry has influenced almost every field of science. Mandelbrot himself has written lots of early papers on the application of fractal mathematics to financial time series. One of them is on the existence of self-similarity in financial time series. Self-similarity is an important statistical property that exists from nature fractal shapes such as tree leaf formations to computer network traffic traces. Here we give the formal definition of statistical self-similarity and we start here to review the construction of MMAR. The definitions will strictly follow Mandelbrot 1997 [A.Fisher, L.Calvet, B.B.Mandelbrot, Multifractality of DEM/$ rates, Cowles Foundation Discussion Paper 1165 (1997)].

\*need to rewrite this part in LaTeX

Definition 1(self-similar measure):

Let X be a subset of the real line, and μ a random measure on X. When

1. for any S ∈*S* and intervals , and are identically distributed whenever , ⊆ X; and



1. for any non-decreasing sequence of interval the random variables  are independent;



then we call μ a self-similar measure on X.

The concept of self-similarity ensures that any section of the data set would show the same statistical properties as any other subset. This is critical when we adopt the classic R/S method to compute our Hurst estimate.

Definition 2(self affinity):

In general, a stochastic process X(t) with X(0) = 0 is called self-affine, if there exists an α such that for any ,



Definition 3(FBM):

A random process X(t) satisfying, for 0<α<1,

1. X(0) = 0 and X(t) is continuous almost surely
2. For any t ≥ 0 and h > 0, the increment X(t+h) – X(t) is normally distributed with mean zero and variance



Is called a Fractional Brownian Motion (FBM) of self-affinity index α.

It can be shown that a stochastic process satisfying this definition exists for each 0<α<1. The Wiener process (or the classic Brownian motion) is an FBM with self-affinity index ½. Recall that, Bacherlier’s model has the variance growing linearly with time, which lies from its independent Gaussian increment assumption.

Definition 4: Let be a fractional Brownian Motion with self-affinity index H, and let be the cumulative distribution function of a self-similar random measure defined on [0,T] ( is a stochastic trading time). Assume furthermore that and are independent.



Define a compound process



Which is related to the price process P(t) by X(t) = ln P(t) – ln P(0). This is the multifractal model for asset returns (MMAR).

Note here the self-affinity index H is the Hurst coefficient we are interested. H generally falls in the range of (0,1) and a larger H means a stronger positive autocorrelation. As a smaller H means a tendency of mean reverting. At H = 0.5, the process has independent increments. We will discuss about it in details in the next section.

\*to be added: how and why MMAR satisfy the three requirements mathematically to be a good model.

Section 3: Hurst coefficient and R/S method

Hurst coefficient, in Mandelbrot’s fractal geometry, is directly related to the fractal dimension, which measures the roughness of a surface. Generally the relationship can be written as D = 2 – H. In this case, a higher Hurst coefficient means a smoother surface. In statistical research, Hurst coefficients have since long been used to measure whether the data is a pure random walk or has long term trends. In finance this has been particularly interesting because the existence of long-term dependency in financial time series has been the hypothesis of many early theories. The efficient market hypothesis prevents the existence of such non-random factor and forces the value of H to be none other than 0.5.

The problems with using H as indicator are: (\* to be expanded and explained)

1. Difficulty to replicate FBM with known H to check the accuracy

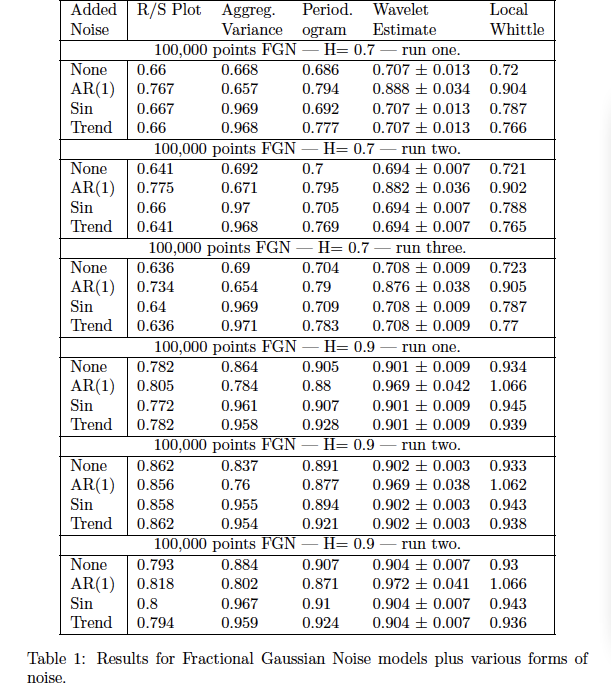
2. There is not a universal or “best” algorithm to measure H in terms of accuracy, measure erros dependent on particular data sets.

3. H varies with the length of the time window for returns. There is not enough financial data so the asymptotic limit behavior suggested by Mandelbrot may never be fully reached.

4. Interpretation of H, is H a meaningful parameter as financial mathematicians have yet to put the use of Hurst coefficient into implying current market parameters successfully?

There are a number of well-known methods to measure Hurst exponent. Most of these methods come from electrical and computer engineering because Hurst exponent has been widely used to measure long-term dependency in signal time series. The time series in economics, however, contains highly non-trivial factors and the non-random trends is mixed with known cycles and irregular spikes and noise related to real world events. That is the reason why Hurst coefficient hasn’t been widely regarded as a standard measurement in finance. The following is a table from a computer science paper giving measurement accuracies of different methods on a sample signal time series [A Practical Guide to Measuring the Hurst Parameter, Richard G. Clegg June 28, 2005 and R. G. Clegg. Statistics of Dynamic Networks. PhD thesis, Dept. of Math.,

Uni. of York., York., 2004.]:



We can see the accuracy of the method strongly depends on the form and the size of the data. For financial time series, the most commonly used method is the classic Rescaled Range Analysis or the R/S Method. An important modified version of this method comes from Lo, who criticized the failure of the previous R/S method in separating short-term dependencies with long-term factors. Interestingly under Lo’s method, the market returns are only governed by the short-term dependence and no long-term dependence was actually presented. Other methods used in finance include aggregated variance, Moody-Wu, etc. However, for our purposes method selection would not necessarily be a problem due to the following reasons.

For our purposes, rather than investigate the efficiency of Hurst estimating. We try to apply H to imply current market parameters. The analyses we are carrying are cross-sectional studies. Since same stock in different time periods give different H value for its return series, it’s hard to directly assign a numerical H value to each and every stock as a label. However, just like the alpha and beta of stocks, we can set a fixed period and do a cross-sectional study among different stocks and portfolios. This way we can compare stocks for the same time window and Hurst coefficient will indeed become meaningful if it exhibits any relationship with return at all. It’s noticed that people have done similar research for hedge fund manager selections. The hedge fund returns are divided into three groups, low Hurst, mid Hurst and large Hurst groups. What the results showed is striking in the sense that high Hurst groups give rather clear higher average returns and it was concluded that people should select high Hurst hedge fund mangers.

R/S method or rescaled range statistic is the original method introduced by Hurst in 1951[H. E. Hurst. Long-term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers, pages 770-808, 1951]. This method was analyzed and named as range over standard deviation or R/S by Mandelbrot and his co-authors in papers published between 1968 and 1977. It’s the most commonly used method in Hurst coefficient estimation. Here we used MATLAB to perform the Rescale Range Calculation.

The R/S statistic is the range of partial sums of deviations of a time series from its mean, rescaled by its standard deviation.

Consider a sample of returns  and let  denote the sample mean 

Then the running sum of accumulated deviation is:



the range,  is the difference between the maximum value of  and the minimum value of  over the time period n.



The standard deviation is then the usual maximum likelihood estimator



Rescaled Range is then defined as:



The problem with this classical R/S statistic is that the mean of the rescaled range may be biased upward by 73 percent as found by Lo. This means the classical rescaled range would have a biased upward to 73 percent for Lo’s example of an AR(1) process. (\*to be added: key difference of Lo’s model in the standard deviation denominator estimation.)

Although aware of the effects of short-range dependence on the rescaled range, Mandelbrot (1972, 1975) did not correct for this bias since the focus was on the relation of the R/S statistic’s logarithm to the logarithm of the sample size as the sample size increases without bound. For short-range dependent time series, the ratio approaches ½ in the limit, but converge to other values when there is a long-range dependence. The ratio is our Hurst exponent and mathematically:

as 

Now H is estimated by calculating the average rescaled range over multiple regions of the data. Beyond some large n, the plot of logarithm of R/S and the logarithm of the sample size n should, as suggested by Mandelbrot, produce a slope that settles down to H. In our case, following the classical rescaled range method we can draw a linear regression line through the set of points, composed of logarithm of R/S and the logarithm of n and record the slope.

Example 1:

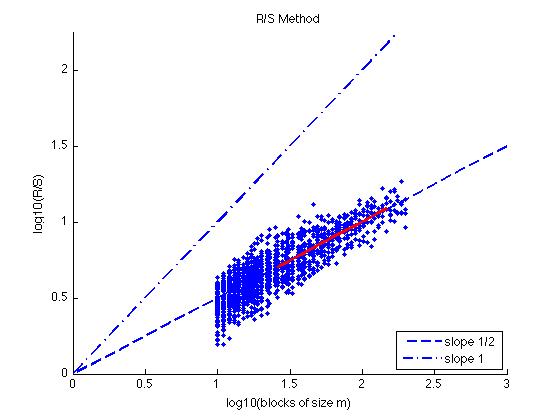
A random time series of 1000 entries should produce a R/S hurst estimate near 0.5

>> data = rand(1,1000);

>> hurst\_estimate(data,'RS',1)

ans =

0.5092



Example 2: a test of Fractional Gaussian Noise sequences generated with Fast Fourier Transform or FFT method.

The fgn070\_S is a simulated FGN sequence with H=0.7 and length=1000;

the fgn080\_L is a simulated FGN sequence with H=0.8 and length=10000.

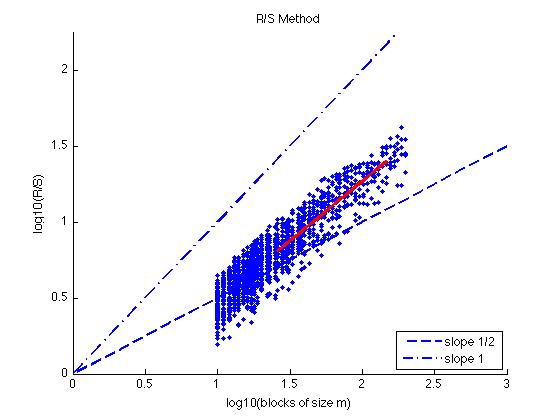
(\*to be added: more details about FFT, in Paxson's "Fast, Approximate Synthesis of Fractional Gaussian Noise for Generating Self-Similar Network Traffic".)

results are:

>> hurst\_estimate(fgn070\_S,'RS',1)

ans =

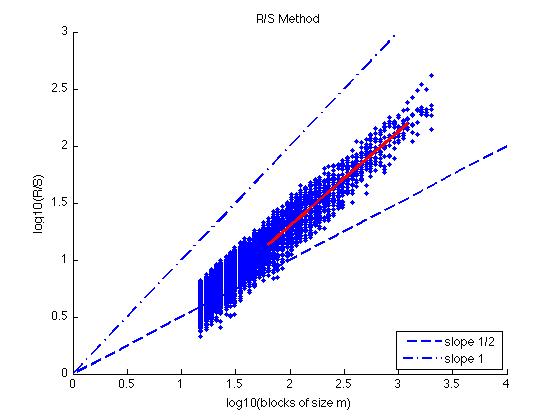
0.7845



and

>> hurst\_estimate(fgn080\_L,'RS',1)

ans =

0.82

Section 4: Hurst coefficients for Kenneth French’s portfolios

The data comes from Kenneth French’s data library, we first test the H values on different industries. We start with the 48 industry daily return data.

Average Value Weighted Returns – Daily of the following 48 industries:

Agric Food Soda Beer Smoke Toys Fun Books Hshld Clths Hlth MedEq Drugs Chems Rubbr Txtls BldMt Cnstr Steel FabPr Mach ElcEq Autos Aero Ships Guns Gold Mines Coal Oil Util Telcm PerSv BusSv Comps Chips LabEq Paper Boxes Trans Whlsl Rtail Meals Banks Insur RlEst Fin Other

\*notes by Kenneth French: This file was created by CMPT\_IND\_RETS\_DAILY using the 200908 CRSP database. It contains value- and equal-weighted returns for 48 industry portfolios.

The portfolios are constructed at the end of June. Datas are from July 1st, 1963 to Aug 31st, 2009.

>>for i=2:49

output(i) = hurst\_estimate(ind(1:11622,i),'RS',0);

end

>> output

output =

Columns 1 through 7

0.8147 0.5493 0.5835 0.5425 0.5420 0.5308 0.5610

Columns 8 through 14

0.5748 0.5858 0.5241 0.6335 0.7201 0.5493 0.5192

Columns 15 through 21

0.5476 0.5914 0.6212 0.5725 0.5903 0.5255 0.5557

Columns 22 through 28

0.5759 0.5459 0.5720 0.5894 0.5873 0.5459 0.5224

Columns 29 through 35

0.5453 0.5761 0.5363 0.5728 0.5626 0.5808 0.5651

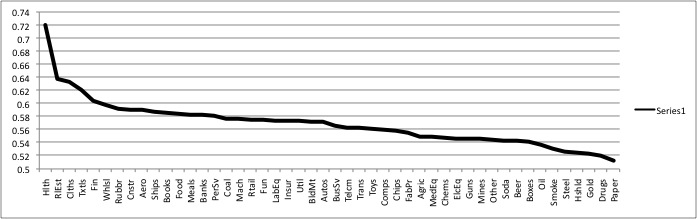
Columns 36 through 42

0.5590 0.5573 0.5738 0.5127 0.5406 0.5621 0.5984

Columns 43 through 49

0.5753 0.5819 0.5818 0.5732 0.6368 0.6042 0.5447

48 Hurst Coefficients are calculated and rearranged into a descending order by its numerical values:



The highest is H= 0.7201, the lowest is H = 0.5192, so every industry is smoother than random walk?

\*to be added: statistics on the Hurst Coefficients, analysis why certain industry present high H

Section 5: test hypothesis that there is relationship between Hurst and average return

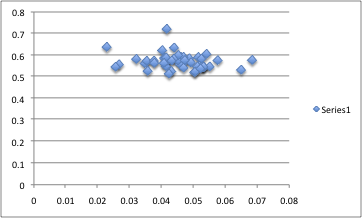
Average returns for every industry has been calculated:

|  |  |  |  |
| --- | --- | --- | --- |
| industry | Hurst Estimate | Return VW | Return EW |
| Agric | 0.5493 | 0.05196868 | 0.077061607 |
| Food | 0.5835 | 0.048742041 | 0.074053519 |
| Soda | 0.5425 | 0.053340217 | 0.081388745 |
| Beer | 0.542 | 0.052961625 | 0.067216486 |
| Smoke | 0.5308 | 0.065014627 | 0.08082602 |
| Toys | 0.561 | 0.034834796 | 0.077882464 |
| Fun | 0.5748 | 0.057665634 | 0.094858888 |
| Books | 0.5858 | 0.040923249 | 0.070321803 |
| Hshld | 0.5241 | 0.043119945 | 0.072769747 |
| Clths | 0.6335 | 0.044013939 | 0.075364825 |
| Hlth | 0.7201 | 0.041694447 | 0.095935497 |
| MedEq | 0.5493 | 0.052817071 | 0.091855102 |
| Drugs | 0.5192 | 0.050401824 | 0.093888315 |
| Chems | 0.5476 | 0.04147651 | 0.075068835 |
| Rubbr | 0.5914 | 0.04600327 | 0.096284633 |
| Txtls | 0.6212 | 0.040289967 | 0.058983824 |
| BldMt | 0.5725 | 0.041852521 | 0.082168301 |
| Cnstr | 0.5903 | 0.046913612 | 0.089276372 |
| Steel | 0.5255 | 0.035736534 | 0.068732576 |
| FabPr | 0.5557 | 0.026739804 | 0.073892617 |
| Mach | 0.5759 | 0.042051282 | 0.079872655 |
| ElcEq | 0.5459 | 0.055150576 | 0.096781965 |
| Autos | 0.572 | 0.035409568 | 0.066620203 |
| Aero | 0.5894 | 0.051564275 | 0.092789537 |
| Ships | 0.5873 | 0.041491137 | 0.061906729 |
| Guns | 0.5459 | 0.051287214 | 0.08293237 |
| Gold | 0.5224 | 0.050584237 | 0.129647221 |
| Mines | 0.5453 | 0.053223197 | 0.100887111 |
| Coal | 0.5761 | 0.068458097 | 0.089751334 |
| Oil | 0.5363 | 0.052269833 | 0.098902082 |
| Util | 0.5728 | 0.037695749 | 0.049844261 |
| Telcm | 0.5626 | 0.037779212 | 0.078584581 |
| PerSv | 0.5808 | 0.032151953 | 0.085789021 |
| BusSv | 0.5651 | 0.049278093 | 0.093284288 |
| Comps | 0.559 | 0.046586646 | 0.08795302 |
| Chips | 0.5573 | 0.0457873 | 0.100185854 |
| LabEq | 0.5738 | 0.045208226 | 0.105710721 |
| Paper | 0.5127 | 0.042383411 | 0.06494063 |
| Boxes | 0.5406 | 0.046972122 | 0.077218207 |
| Trans | 0.5621 | 0.040813973 | 0.072625194 |
| Whlsl | 0.5984 | 0.045313199 | 0.084297023 |
| Rtail | 0.5753 | 0.047338668 | 0.078391843 |
| Meals | 0.5819 | 0.052560661 | 0.082825675 |
| Banks | 0.5818 | 0.043898641 | 0.076528136 |
| Insur | 0.5732 | 0.043235244 | 0.072631217 |
| RlEst | 0.6368 | 0.02289365 | 0.083358286 |
| Fin | 0.6042 | 0.054149028 | 0.092160558 |
| Other | 0.5447 | 0.025660816 | 0.078288591 |

No simple obvious relations were spotted:

y-axis: portfolio return,

x-axis: hurst coefficient



x-axis: portfolio return,

y-axis: hurst coefficient

log-log plot wouldn’t give any immediate picture either.

\*to be tested:

* use H from different subdivisions of time period
* try other portfolios
* try to fit with exponents and other possible functional relations
* different method to estimate H
* try to examine against market parameters other than returns
* try volumes, volatilities
* try monthly data

Section 6: Related studies

\*controversial debates and arguments about MMAR and H interpretation. Intuition and logics behind the model.

Section 7: Conclusion

Section 8: References