

Rescaled Range(R/S) analysis of the stock market returns

Prashanta Kharel,
THE UNIVERSITY OF THE SOUTH*

29 Aug, 2010

Abstract

The use of random walk/ Gaussian distribution to model financial markets is a well established practice. However, this paper attempts to verify Mandelbrot's claim¹ that such models are not good models as markets have persistence, bias and fluctuate way more than Gaussian statistics would allow. Furthermore, using Matlab[©] based rescaled range(R/S) analysis, the paper aims to understand the long memory effects, fractal statistical structure, and the presence of cycles, especially in the markets of returns.

Keywords: Time series analysis, Fractal Market Hypothesis, non-Gaussian statistic, R/S analysis.

1 Introduction

The traditional mathematical model for financial markets, following Louis Bachelier's suggestion, builds upon Gaussian statistic partly because such statistic presents a wide range of mathematical techniques at one's disposal. Furthermore, the statistic provides a good approximation to market behaviour as long as the markets are calm. However, it is the during the crucial periods of high volatility that the traditional Gaussian statistic deviates from the real happenings of the market. The stock market crash of September, 1989, the dot-com bubble of late 1990s, and even the recent housing market crash of 2007-2008 are examples of such periods with high volatilities, which are significantly more frequent than that predicted by the traditional Gaussian statistic.

An interesting comparison with highway driving underscores the importance

*Thanks Dr. Benito T. Szapiro for all the guidance, and the donor for the financial support

¹The (mis)Behavior of Markets: A fractal view of Risk, Ruin, and Reward by *Benoit B. Mandelbrot and Richard L. Hudson*

of understanding such periods with high volatilities. It is a known fact that there are many more cars than trucks in a highway. However, if one plans his car safety and driving habits just based on the low probability of crashing against a truck, he will not survive long, as this one event will wipe him out of this earth. Such rare events have devastating consequences, and the same is true for financial markets.

The traditional mathematical models for financial markets are based on two critical assumptions: the price change follows a Gaussian distribution, and the price changes are independent. However, using a systematic mathematical treatment of standard time series of stock market returns, this paper will first verify Mandelbrot's claim that the random walk/Gaussian models are not good estimates of a real market behaviour. Furthermore, we investigate the long term dependence, fractal statistical structure, and presence of cycles, to support our claim that markets are persistence and biased.

2 Goodness of fit of real time series with normal distribution

Let x be a time series, where $x = x_1, x_2, \dots, x_n$. The time index is unimportant in general. Now we detrend the time series x to remove any short memory effect. Removing a continuous, piecewise linear trend from the series x , where each linear segment is five points long, produces the detrended time series x_d . Then, we calculate the differences between adjacent elements of the detrended series, x_d , to obtain a differenced series, x_{diff} . This differenced series is rescaled by subtracting it by the mean value, x_m , and dividing it by the standard deviation, s_n , of the time series x . Here, x_m and s_n is defined as:

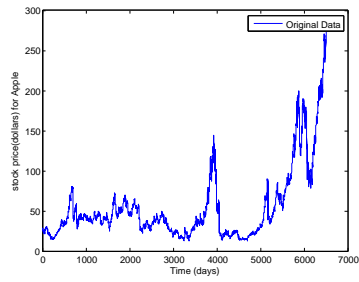
$$x_m = \frac{x_1 + x_2 + \dots + x_n}{n}, \text{ and } s_n = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - x_m)^2 \right)^{\frac{1}{2}}.$$

This rescaled series, x_r , has a mean of zero. We plot x_r together with $\pm 3\sigma$ bounds. Finally, we compare the rescaled series, x_r , with normal distribution to see the goodness of fit.

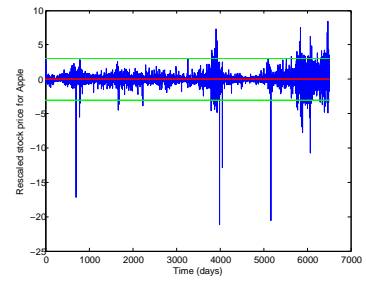
In our analysis, we started with a time series of daily closing stock market returns for Apple Inc.² covering a period from Sep, 1984 to Aug, 2010. The data spanned a time period of more than three decades, which we assumed was a suitable data range for our analysis. Matlab[©] based analysis produced the following series produced the results in fig:1.

It is apparent from the results, especially from figure c, that the rescaled data deviates significantly from the normal distribution at the tails of the distribution. Furthermore, one can see significant amounts of period of high volatility in figure b, where the price changes swings more than $\pm 3\sigma$

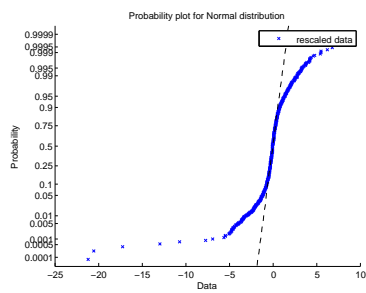
²the historical prices for Apple Inc.(AAPL) can be downloaded from the Yahoo! finance



(a) Stock Prices for Apple Inc.



(b) Rescaled Stock Prices



(c) Stock Prices for Apple Inc.

Figure 1: Apple Inc.(AAPL)

more frequently than that predicted by my Gaussian distribution. One can also notice the recurring periods of high volatility followed by periods of relative calmness, contradictory to one predicted by Gaussian statistic. The Gaussian distribution estimates the probability of price changes being 15σ s away from the mean is approximately 10^{-9} . However, this is obviously not the case for the Apple stock price change. Just in a period of 36 years, the price changes exceeded this limit around three times. Apparently, “rare” events are not that rare in real life.

Analysis of Dow Jones Industrial average (DJI) daily closing prices covering a period from Oct 1928 to Aug 2010 produced similar results. The graph comparing the rescaled time series for this data with normal distribution is presented in fig:2.

3 Rescaled range analysis

Now that we have verified that Gaussian statistic is not a good statistic to model financial markets, we consider an alternative analysis: rescaled range (R/S) analysis.

- 1 Let x be a time series, where $x = x_1, x_2, \dots, x_m$. First we convert the time series into a logarithmic ratios of length $n=m-1$:

$$N_i = \log\left(\frac{x_{m+i}}{x_i}\right), i = 1, 2, 3, \dots, m-1$$

- 2 Now we divide this time period into A contiguous subperiods of length k , such that $A \cdot k = n$. Each subperiod S_a , $a = 1, 2, 3, \dots, A$ has an element $N_{j,a}$, $j = 1, 2, 3, \dots, k$. For each S_a of length k the average value is e_a , where

$$e_a = \frac{1}{k} \sum_{j=1}^k N_{j,a}$$

- 3 Then we calculate the time series of cumulative differences, $X_{j,a}$, from the mean value for each subperiod S_a is:

$$X_{j,a} = \sum_{i=1}^j (N_{i,a} - e_a), j = 1, 2, 3, \dots, k.$$

- 4 The range within each subperiod S_a is then defined as:

$$R_{S_a} = \max(X_{j,a}) - \min(X_{j,a}) \text{ where } 1 \leq j \leq k.$$

- 5 Then we calculate the sample standard deviation for each subperiod S_a where:

$$S_{I_a} = \left(\frac{1}{n} \sum_{j=1}^k (N_{j,a} - e_a)^2\right)^{\frac{1}{2}}$$

Table 1: S&P 500

Returns(in days)	R/S-Hurst	$E(\frac{R}{S})$	stdev
20	0.577	0.554	0.0378
5	0.558	0.544	0.0183
1	0.535	0.533	0.0081

Table 2: Dow Jones Industrial average

Returns(in days)	R/S-Hurst	$E(\frac{R}{S})$	stdev
20	0.565	0.550	0.0316
5	0.568	0.538	0.0156
1	0.552	0.529	0.0070

- 6 Therefore, the rescaled range for each subperiod S_a is equal to $\frac{R_{S_a}}{S_{S_a}}$. Hence, the average value of R/s for length of k is defined as:

$$(R/S)_k = \frac{1}{A} \sum_{a=1}^A \left(\frac{R_{S_a}}{S_{S_a}} \right)$$

- 7 We repeat the steps 1 through 6 by increasing the value of k such that $\frac{m-1}{k}$ is an integer value until $k = \frac{m-1}{2}$. We plot $\log(R/S)_n$ as a function of $\log(n)$ and use least squares regression to estimate the slope, which is the Hurst exponent H .

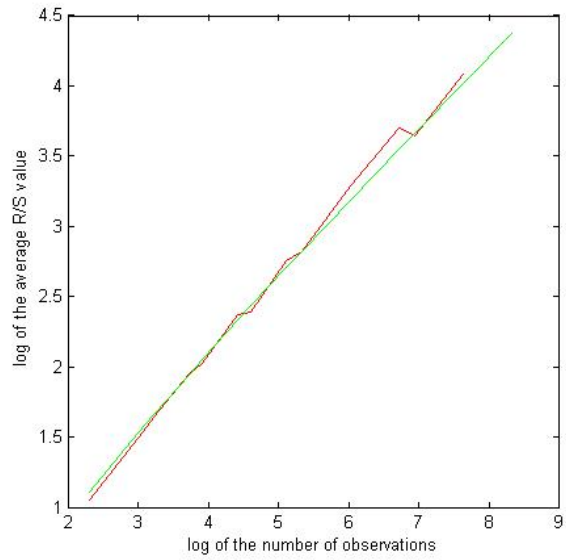
- 8 We also define a test statistic, V , which is: $V_k = \frac{(R/S)_k}{\sqrt{k}}$

A matlab algorithm was written to calculate the rescaled range and Hurst exponent, H . Furthermore, to understand the significance of the results that we get from this algorithm, we wrote a different algorithm based on monte carlo simulation to calculate the expected value of R/S . Typically $H=0.5$ implies an independent process. $0.5 < H \leq 1$ implies persistent time series that have long memory effects. Whereas, $0 \leq H < 0.5$ implies antipersistence, covering less distance than an independent process.

We considered three different time series data- S&P 500 Index, RTH(Jan 1950 - Aug 2010), Dow Jones Industrial Average(Oct 1928-Aug 2010), and DAX (Nov 1990- Aug 2010)- for our analysis. Figure 2 is an illustration of a plot of $\log(\frac{R}{S})_n$ vs $\log(n)$ for Dow Jones Industrial closing prices.

Refer to the table for the results of R/S analysis.

The analysis of each time series involves detrending the data using AutoRegressive AR(1) residuals. This removes the short term memory process so that long memory effects can be investigated.



(a) Dow Jones RS analysis

Figure 2: Dow Jones RS analysis

Table 3: DAX

Returns(in days)	R/S-Hurst	$E(\frac{R}{S})$	stdev
20	0.604	0.538	0.0146
5	0.621	0.554	0.0333
1	0.652	0.567	0.0707

The analysis shows that all three markets show some form of persistence because the Hurst value is above 0.50. However, German stock market(DAX) generally showed more persistence than the US stock market. Particularly significant was the Hurst exponent for 20 days returns of DAX from 1990 to 2010. The Hurst exponent was 4.5 standard deviations away from the expected value, which was calculated using Monte Carlo Simulations. Clearly, there is long memory process at work in these markets. Furthermore, as the sample size increases(daily returns has more data points than weekly returns) the vale of R/S analysis tends to 0.50.

4 Conclusion

In this paper, we showed that the traditional Gaussian model is not a good model for financial markets because the distribution of the price changes for real data are more flat-tailed than normal distribution. Furthermore, using R/S analysis and the values of Hurst exponents for the markets, we showed that markets have persistence, bias ans demonstrate long memory effects. Although Peter's[1991] suggests a four year cycle for Dow Jones Industrials (1888 - 1990), we did not find evidence of any such cycle. Since R/S analysis depends more on the time frame than on the number of data points, it is possible that we were unable to find the cycles because our time frame was different than Peters.Also, a paper by Lo[1991] suggests that using Auto-correlation to estimate H presents a problem, as the regression coefficients are biased.

A next step would be to use modified R/S analysis to calculate H without any bias and ultimately to compare real time series with fractal series.

References

- [1] Edgar E. Peters, "Fractal Market Analysis", John Wiley & Sons, Inc, 1994.
- [2] A.W. Lo, "Long-Term Memory in Stock Market Prices", *Econometrica*, 1991.
- [3] Edgar E. Peters, "Chaos and Order in the Capital Markets", New York, John Wiley & Sons, Inc, 1991.
- [4] Benoit T. Mandelbrot and Richard L. Hudson, "The (mis)Behavior of Markets: A fractal view of risk, ruin and reward.", Basic Books, 2004.