Extending the Variance Ratio Test
to Visualise Structure in Data:
An Application to the S&P 100 Index

by

Andreas Lindemann[*,**]
Christian L. Dunis[*,**]
Paulo Lisboa[*]

(*Liverpool John Moores University and **CIBEF)

September 2004

Abstract

The aim of this paper is to present a method that is able to graphically describe the amount of structure in a time series. In the following, we define as 'structure' the extent to which a time series is either trending or mean-reverting (that is showing pockets of positive as well as negative autocorrelation). What we define as being trending, respectively mean-reverting, should be seen in relation to the characteristics of a random walk.

Testing most of the constituents of the Standard & Poor's 100 index for structure and using a modified variance ratio that focuses on the whole ratio profile rather than an individual ratio, we detect trending as well as mean-reverting structure over a time period of more than 10 years.

Keywords

Efficient Market Hypothesis (EMH), Random Walk, Variance Ratio Statistics (VR), Variance Ratio Profile.

_____

Christian Dunis *is Professor of Banking and Finance at Liverpool John Moores University and Director of CIBEF (E-mail: cdunis@totalise.co.uk).*
Paulo Lisboa *is Professor in Industrial Mathematics at Liverpool John Moores University and Head of the University's Graduate School (E-mail: p.j.lisboa@livjm.ac.uk).*
Andreas Lindemann *is an Associate Researcher with CIBEF (E-mail: afealind@livjm.ac.uk) and currently working on his PhD thesis at Liverpool Business School.*
CIBEF *– Centre for International Banking, Economics and Finance, JMU, John Foster Building, 98 Mount Pleasant, Liverpool L3 5UZ.*

1. INTRODUCTION

The aim of this paper is to present a method that is able to graphically describe the amount of structure in a time series. In the following, we define as 'structure' the extent to which a time series is either trending or mean-reverting (that is showing pockets of positive as well as negative autocorrelation). What we define as being trending respectively mean-reverting should be seen in relation to the characteristics of a random walk.

The motivation for the paper is derived from the fact that for modelling financial time series and developing quantitative tools for trading, it is necessary to know as much as possible about the structure of the time series. Therefore a tool visualising structure in data could be quite useful in the above-mentioned tasks.

In section 2, we give a brief overview of the role that variance ratio tests play in the financial literature. In section 3, we introduce and describe artificially created time series portfolios, which help us to assess the variance ratio method in a controlled environment. Section 4 presents the variance ratio and its results when applied to these artificial data. The same section reveals some problems with this method when it comes to trending time series. Section 5 presents the idea of a modified variance ratio and shows its results when applied to the same data sets. Section 6 examines time series of stocks of the Standard & Poor's 100 index constituents using the modified method, while the paper concludes with some final remarks in section 7.

2. LITERATURE REVIEW

A lot of effort has been made to examine the efficiency of market price formation[1]. Most of the work has been centred around the Efficient Market Hypothesis (EMH) which at its heart states that prices follow a random walk.

One implication of a random walk is that returns are not predictable. This has been extensively tested by judging the value of several forecasting methods such as technical analysis. For example see Levich and Thomas (1993), Osler and Chang (1995), Neely *et al.* (1997) for statistically significant excess returns using technical rules on the USD exchanges rates and Brock *et al.* (1992) and Sullivan *et al.* (1998) for stock markets.

Another implication of a random walk is of statistical nature and states that the variance of a sample must be linearly related to the sampling interval. In other words, the variance of asset returns is directly proportional to the holding period. This property can and has been tested by what is known as the variance ratio tests. First amongst these have been Lo and MacKinlay (1988) who developed a test statistics based on that idea. Their paper concludes with the finding that stock returns do not follow random walks. Other applications of the variance ratio test can be found for example in Chow and Denning (1993), Richardson (1993) and Smith and Ryoo (2003) for stocks as well as Liu and He (1991) and Fong *et al.* (1997) on foreign exchange rates.

Chow and Denning's (1993) results are based on the multiple variance ratio (MVR) test where the random walk hypothesis is rejected if any one of a set of different order variance ratios differs significantly from the test statistics. Groenendijk *et al.* (1998) have taken the idea of the variance ratio one step further. They generalise the variance ratio equation to investigate what can be learned from using other moments in the ratio statistics besides

---

[1] Interesting surveys are provided by Fama (1970,1991), Pagan (1996) and Campbell *et al.* (1997).

the variance. Burgess (1999) provides another expansion by looking at the joint distribution of the set of variance ratio statistics (variance ratio profiles). This expansion leads to better results than a single ratio alone: admittedly, the value of an individual variance ratio can be affected by the fact that short term positive autocorrelation combined with long term negative autocorrelation may cancel out.

Our study will make use of that extension transforming the results into a graphically user-friendly plot.

3. CREATING ARTIFICIAL DATA FOR A CONTROLLED ENVIRONMENT

In the following, we create a number of artificial time series. Some are based on a random walk, some are trending and some are mean reverting. The use of those series lies in the possibility to measure under controlled circumstances the ability of methods claiming to be able to discriminate structure from random walks.

Each type of time series is represented by a portfolio of 100 artificially created series. Each series consists of 500 values (interpreted in the following as daily percentage changes or simply returns). The 'random' portfolio was created by sampling the daily changes from a normal distribution (N(0,1)). The 'trending' series were generated by sampling from a normal distribution, this time with a mean different from zero (N(0.15,1)). The creation of the 'mean reverting' series has been done by also sampling the returns from a normal distribution (N(0,1)), but in addition subtracting 0.2% when the return was positive and added 0.2% when the return was negative.

The figures below show an overview of the different time series portfolios we are using. The x-axis represents time while the y-axis defines the levels of the series[2].
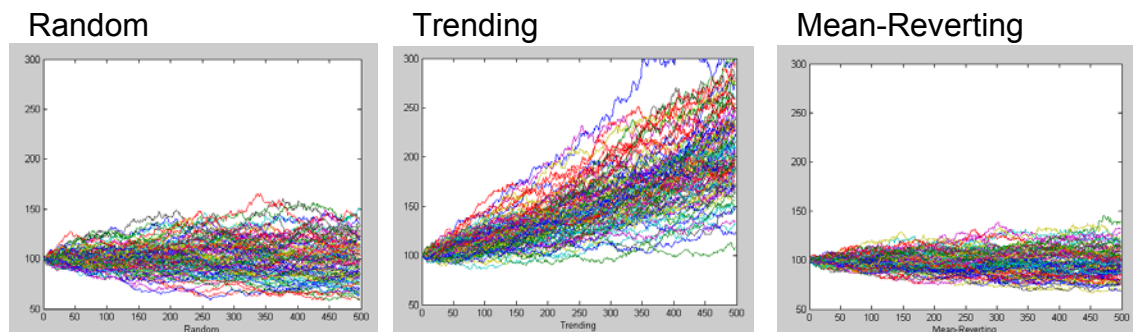

Fig. 1: Artificial time series.

4. THE VARIANCE RATIO

4.1. Variance Ratio Definition

The variance ratio is a very well studied method used to test the random walk hypothesis. Doing so, the Variance Ratio (*VR*) is built around the 'linear property' that the increments of a random walk must be a linear function of the time interval[3].

---

[2] The transformation from a series of returns to a series of levels has been done by using the cumulative product rather than the cumulative sum of the returns.
[3] E.g. using log returns, the variance of r(t) + r(t-1) must be twice the variance of r(t) (see Campbell *et al.*(1997), p. 48).

Under the assumption of a random walk with uncorrelated daily changes Δy_t, the variance of these changes increases linearly with the time interval so that:

$$VAR(\Delta^\tau y_t) = \tau VAR(\Delta y_t)$$ [1]

where:

τ defines the length of the long term variance
y_t is the actual time series in levels
Δy_t is the daily change $y_{t+1} - y_t$
$\Delta^\tau y_t$ is the long term change $y_{t+\tau} - y_t$

Then, *VR* is defined as the ratio of the long term variance (calculated over period τ) normalised by its single-period variance:

$$VR(\tau) = \frac{\sum_t \left(\Delta^\tau y_t - \overline{\Delta^\tau y} \right)^2}{\tau \sum_t \left(\Delta y_t - \overline{\Delta y} \right)^2}$$ [2]

where:

$\overline{\Delta y}$ is the mean value of the short term changes

$\overline{\Delta^\tau y}$ is the mean value of the long term changes.

Under the null hypothesis of a random walk we have $VR(\tau) = 1$.

There is a strong relation between *VR* and autocorrelation. Starting with *VR*(2), Lo and MacKinlay (1988) show that in this case, the ratio can be rewritten as:

$$VR(2) = 1 + \rho(1)$$ [3]

where: $\rho(1)$ defines the first-order autocorrelation coefficient.

Since the autocorrelation is zero in the case of a random walk, *VR*(2) should simply be 1. It has been shown that higher order *VR*(ρ) are specific combinations of the first k-1 autocorrelation coefficients of y_t, with linear declining weights (Campbell *et al.*, 1997).

4.2. Variance Ratio Profiles

VRs of different orders can be separately used to check the amount of autocorrelation in a time series and therefore help determine whether it is trending or mean-reverting. Depending on the choice of the variance ratio order, one looks at a specific autocorrelation and therefore on a specific definition of trends (i.e. short term, middle term or long term).

Since in this case the judgement of a series being trending or not is depending on the specific choice of the *VR* order, it seems best to look at the whole profile rather than picking one specific order. Especially when there is the possibility that short term positive autocorrelation combined with long term negative autocorrelation may cancel each other out when looking only at a specific *VR* order.

This can be done by calculating the *VR* for different lengths of τ resulting in a *VR* profile of that specific time series. By comparing the profile of a time series with those profiles of artificially trending and/or mean-reverting series, one is then able to get a good feeling for the structure of the series under investigation.

For our three artificial time series portfolios, the profiles using a long term variance ratio up to 20 days are shown below where the x-axis specifies the τ of the *VR* and the y-axis shows the *VR* value.
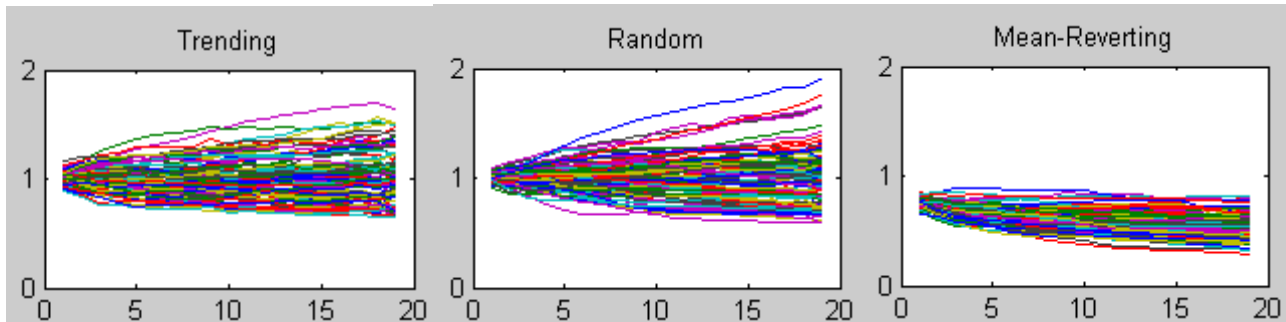

Fig 2: Variance ratio profiles (for τ =2…20)
of the time series of the artificial portfolios.

4.3. Variance Ratio Plots

Unfortunately, using profiles alone does not seem to help in classifying whether a time series as trending, random or mean-reverting. To do so, it would be helpful to find a way to plot the profile of a series in a graph as a single point. One way to achieve this is by defining archetypes of profiles. The extent to which a profile matches one of the archetype profiles could be used as a coordinate in 'archetype space' (each axis in that space represents the extent of similarity to one specific archetype profile).

A useful tool to define those 'archetype profiles' is the standard technique of principal component analysis (PCA)[4] since this method identifies the orthogonal (i.e. uncorrelated) factors explaining the variance of the underlying time series. In our case, PCA is applied to the co-variance matrix of random walk profiles.

These factors, the eigenvectors of this co-variance matrix, can then be used as our 'archetypes'.

The first 5 eigenvectors are shown below (starting from top to bottom).

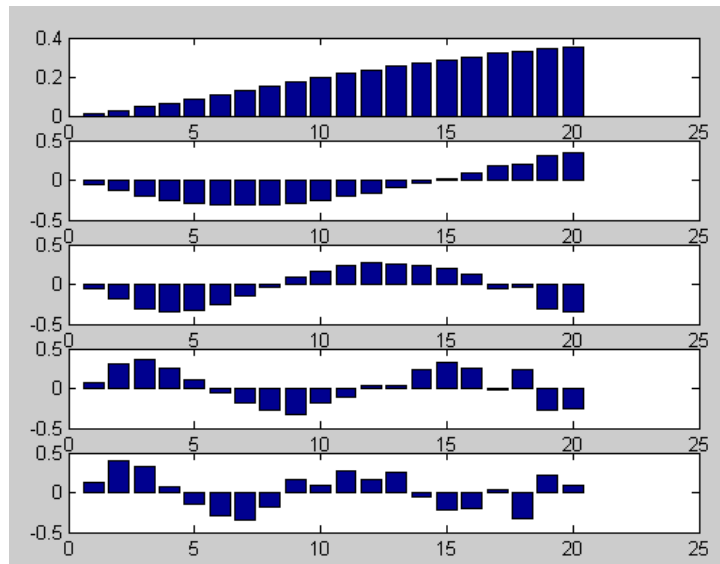[4] For more details on PCA, see Jolliffe (1986).

Fig 3: The first 5 eigenvectors (archetypes) as a result of the application
of PCA analysis to the co-variance matrix of random walk profiles.

As it turns out, the first eigenvector explains as much as 93% (which is the eigenvalue expressed in percentage terms) of the variance of the co-variance matrix, while the second explains 4.4% and the third 1.3%. This result allows us to restrict ourselves to the usage of the first 2 or 3 eigenvectors (respectively archetypes) while still being able to cover most of the structure. Using these few vectors enables us to plot our results into either 2 or 3 dimensional space.

Defining the similarity of the actual profile with one of the archetypes is achieved by a simple vector multiplication [(profile of actual time series) *x* (eigenvector)], and we are now able to characterise a profile by just two values: its similarity to the first and second eigenvectors.

Applying the above described method to our artificial data portfolios, we are now able to compare the results of different time series behaviour (i.e. trending and mean-reverting) to each other.

The figure below analyses the artificial portfolios where figure 4a presents the trending, figure 4b the random and figure 4c the mean-reverting time series portfolios.

The first box of each figure gives an overview of all the profiles while the second box presents the average profile of the whole portfolio. The interesting plots are in box three and four. Here we finally get time series plotted as single points where the coordinates are dependant on the 'similarity' to the first two archetypes.

A single dot in the plot in box three represents a time series of one of our artificial portfolios while the square indicates the centre of that data cloud. Box four is added for reasons of comparability and is based on random time series whose centre is indicated by a triangle. The scales of plots in box three and four are kept the same to simplify the comparability.
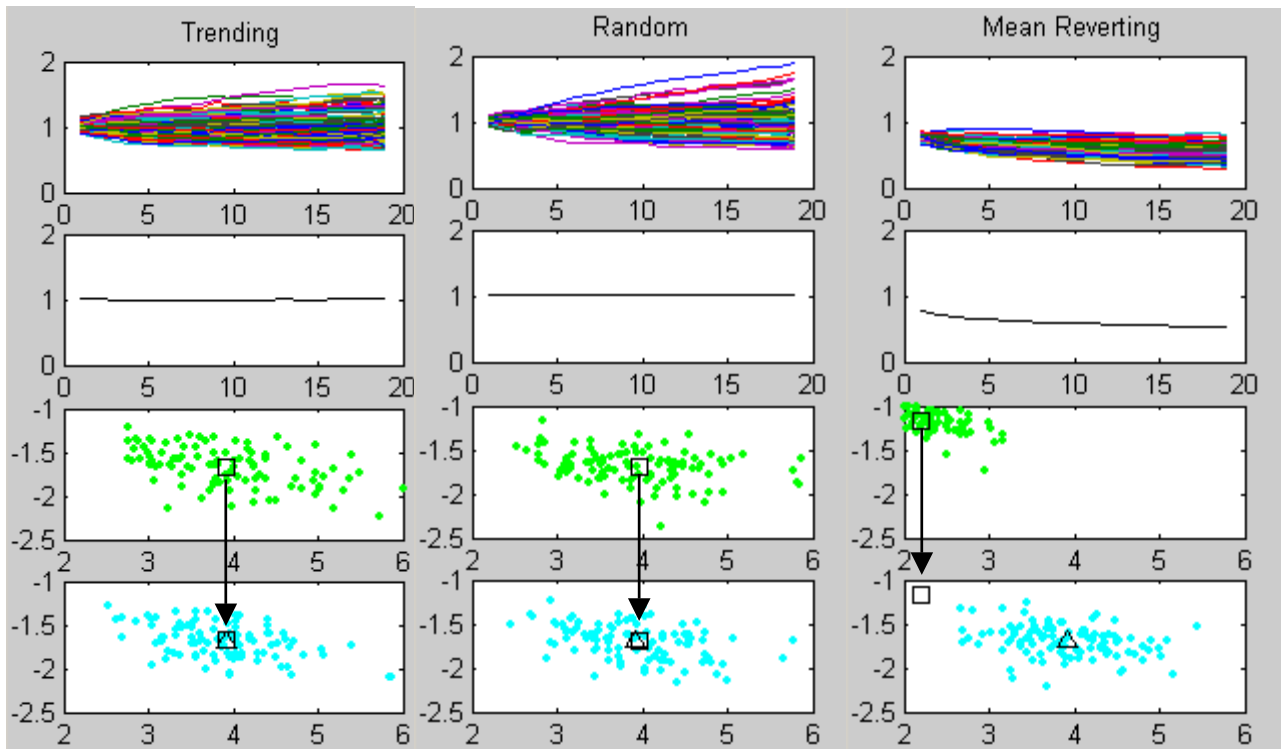
Fig. 4a   Fig. 4b   Fig. 4c

The artificial time series are mapped as spots into two dimensional space
and compared to a random walk series.

As we see on the graphs above, the mean-reverting time series can easily be distinguished from the random series by looking at their centres in the third boxes. Unfortunately the trending series pretty much looks the same like the random walk. This is a problem of the *VR,* which we are going to look closer at in the next section.

4.4. Variance Ratio Problem

Using the above definition of *VR* has one big shortcoming. It is not able to properly discriminate between trending and random series.

To find out why, we define two trending time series. One is created by sampling from a normal distribution N(0,1) where the returns are rearranged in a way that we have an up-trend at the beginning and a down-trend at the end. The histogram of that series will look like a normal distribution but due to the rearrangements, it will show pockets of positive autocorrelation at the beginning as well as at the end. The second time series is also sampled by a normal distribution N(0,1) and then a small value is added to each return (e.g. 0.15%). The histogram of that series will look like the density function of a normal distribution N(0.15,1). As a result, we have a trending time series with mean zero (the first series) and a second trending series with non-zero mean. The inherent problem with the *VR* is now that we have to normalise the nominator as well as the denominator by subtracting the mean. As a result, this will not affect series one but series two is suddenly transformed into a normal random walk [N(0,1)]. The reason why this affects only trending time series lies in the fact that the mean-reverting time series already has a mean close to zero.

The *VR* is able to find trends generated by a 'structured' sampling from a N(0,1) but it is blind to a trend generated from a N(0.15,1). This is the reason why the data cloud in figure 4a and figure 4b look the same.

5) THE MODIFIED VARIANCE RATIO

5.1. Modified Variance Ratio Definition

The easiest way to avoid this problem is simply by leaving out the subtraction of the mean value in the *VR* equation. This will change the original *VR* equation [2] to the modified equation below:

$$VR^*(\tau) = \frac{\sum_t \left(\Delta^\tau y_t \right)^2}{\tau \sum_t \left(\Delta y_t \right)^2} \quad\quad\quad [4]$$

5.2. Modified Variance Ratio Plots

Using the modified *VR** and using the same artificial time series as before, we should now be able to discriminate between trending and random time series.

Fig. 5a   Fig. 5b   Fig. 5c
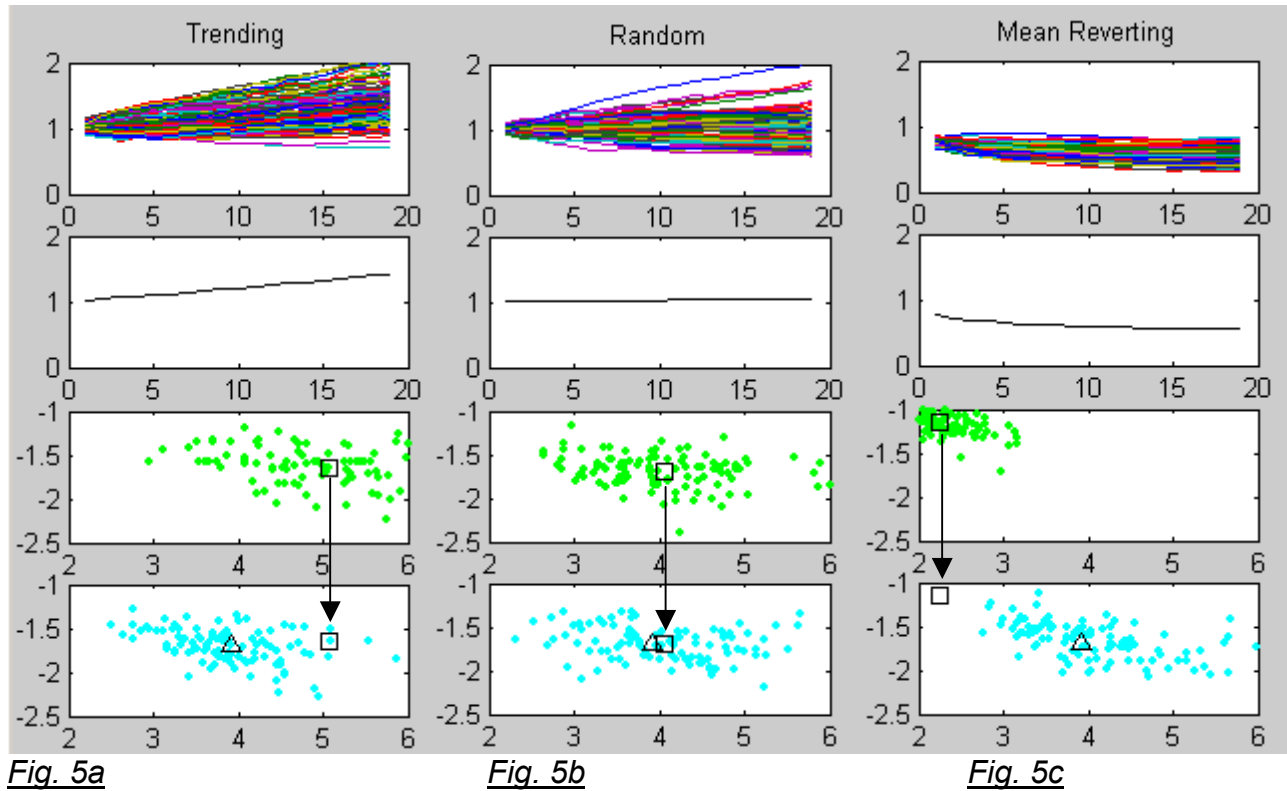
Mapping the artificial time series into two dimensional space
using the modified VR equation [4] to calculate the profiles.

Figure 5 clearly shows that the modified *VR* is able to distinguish between random and trending series. First of all the average profile of the trending series is no longer flat anymore but up-trending and, secondly, the mean-reverting series are plotted to the left side, while trending series are more to the right side.

6) APPLYING THE MODIFIED VARIANCE RATIO TO THE S&P100 INDEX

We have shown that the modified *VR* method is a useful tool to test time series for structure on artificial time series. The next step is to test real financial data for structure. The standard variance ratio is a common tool in the literature to test financial data for the random walk hypothesis most often resulting in the acceptance of this hypothesis. It would now be interesting not only to use the modified *VR* equation but also to use the whole profile rather than a single pre-defined *VR*. We apply our methodology to the constituents of the Standard & Poor's 100 index. All time series are obtained from an historical database provided by Datastream. The time period considered is 1980 to 2002.

Total dataset	Days	Beginning	End
	6000	01 January 1980	31 December 2002

Table 1: The EUR/USD dataset

This long time period reduces the time series available from 100 to 69. These remaining companies are:

Number	Variable / Company	Number	Variable / Company
1	3M	35	HALLIBURTON
2	ALCOA	36	HEINZ HJ
3	ALTRIA GP.	37	HEWLETT-PACKARD
4	AMER.ELEC.PWR.	38	INTEL
5	AMERICAN EXPRESS	39	INTL.BUS.MACH.
6	AMERICAN INTL.GP.	40	INTL.PAPER
7	ANHEUSER-BUSCH COS.	41	JOHNSON & JOHNSON
8	AT & T	42	JP MORGAN CHASE &.CO.
9	AVON PRODUCTS	43	LIMITED BRANDS
10	BANK OF AMERICA	44	MAY DEPT.STORES
11	BANK ONE	45	MCDONALDS
12	BAXTER INTL.	46	MEDTRONIC
13	BLACK &.DECKER	47	MERCK &.CO.
14	BOEING	48	MERRILL LYNCH &.CO.
15	BOISE CASCADE	49	NATIONAL SEMICON.
16	BRISTOL MYERS SQUIBB	50	PEPSICO
17	BURL.NTHN.SANTA FE C	51	PFIZER
18	CAMPBELL SOUP	52	PROCTER & GAMBLE
19	COCA COLA	53	RADIOSHACK
20	COLGATE-PALM.	54	RAYTHEON 'B'
21	COMPUTER SCIS.	55	ROCKWELL AUTOMATION
22	DELTA AIR LINES	56	SARA LEE
23	DOW CHEMICALS	57	SCHLUMBERGER
24	DU PONT E I DE NEMOURS	58	SEARS ROEBUCK &.CO.
25	EASTMAN KODAK	59	SOUTHERN
26	ENTERGY	60	TEXAS INSTS.
27	EXELON	61	TOYS R US HOLDINGS CO.
28	EXXON MOBIL	62	UNISYS
29	FEDEX	63	UNITED TECHNOLOGIES
30	FORD MOTOR	64	WAL MART STORES
31	GEN.DYNAMICS	65	WALT DISNEY
32	GENERAL ELECTRIC	66	WELLS FARGO &.CO
33	GENERAL MOTORS	67	WEYERHAEUSER
34	GILLETTE	68	WILLIAMS COS.
		69	XEROX

Table 2: Components of training dataset

The profiles, i.e. the eigenvector plot for this portfolio, are presented below. It is split up into the first 3000 data points and the remaining 3000 data points. In order to adjust for the different characteristics of each time series regarding its standard deviation, we have smoothed each time series by using a rolling window of 50 days length to normalise the series (by dividing the actual value by the standard deviation of the last 50 days).

Here again, the time series in question are represented by spots in box three while additional random series are added as spots in box four in order to provide a benchmark for the inherent structure in the investigated series. As it turns out, during the first sub-period, the SP100 portfolio was centred around the same mean value as the random portfolio. But unlike the random benchmark, the SP100 portfolio revealed several more extreme values reflecting both trending and mean-reverting behaviour. While during the second half of our investigation period, the series have shown a strong mean-reverting behaviour, even moving the whole centre of the portfolio cloud towards the mean-reverting (left) side of the plot. Looking at the whole period length of 6000 days, the overall mean-reverting effect clearly dominates the slightly trending behaviour discovered in the first half of our period (as shown in figure 6a).
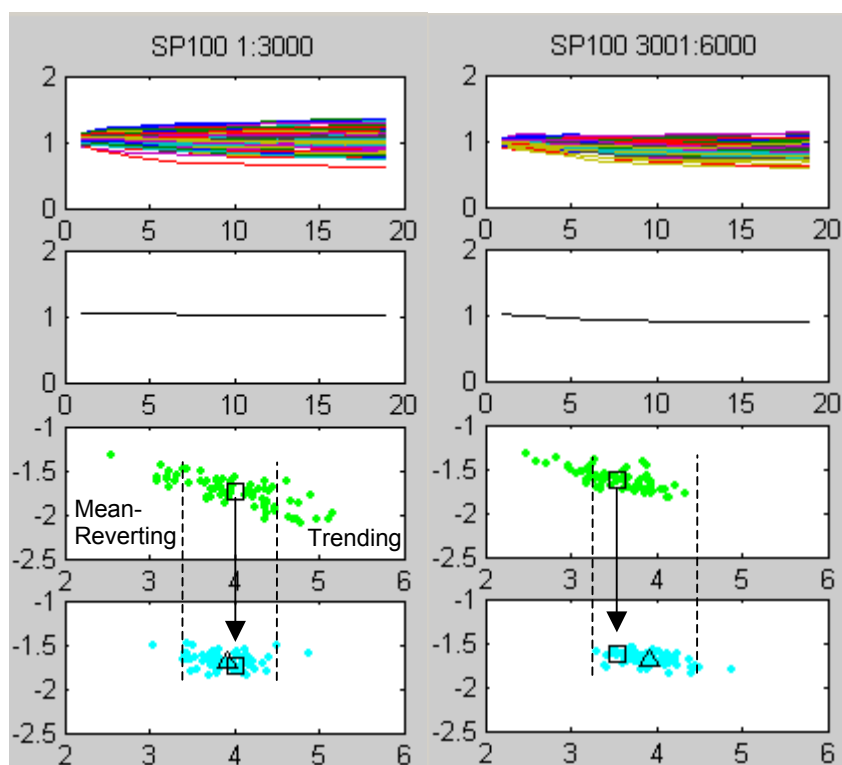

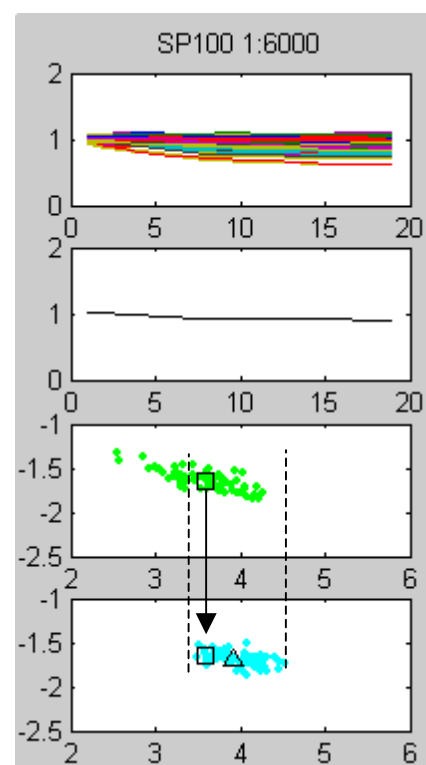*Fig. 6a*   *Fig. 6b*   *Fig. 6c*

Plotting the S&P constituents into two dimensional space and
comparing them with a random walk portfolio.

Using a data length of 1000 days for each sub-period and calculating the centre of the SP100 cloud as well as for a random portfolio using the same number of time series of the same length, one can plot the movement of those centres as shown below. The left graph shows the centre movement of the SP100 portfolio while the right graph shows the result for a random walk portfolio. The SP100 obviously exhibits stronger movements than could be expected from a random walk. The ellipse marks the space where the centres of a random walk portfolio could be expected with a 90% probability. The triangle shows the centre of the ellipse.
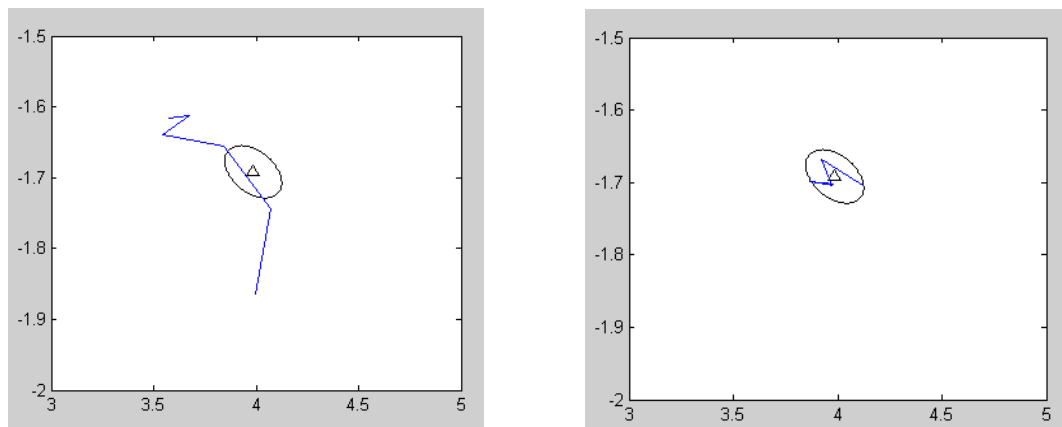
Fig. 7: Tracking the movements of the centre of the S&P constituents cloud [left figure] compared to the movements of a random walk portfolio [right figure].

The evolution of the whole SP100 portfolio cloud is analysed in more detail in the appendix, where the step length is 1000 days rather than 3000. In addition three companies have been randomly picked and followed over the total length of investigation. They are marked using a star, a diamond and a square.

7. CONCLUDING REMARKS

We have shown that the Variance Ratio can be a useful tool to detect structure (trending and/or mean-reverting behaviour) in time series. By using the whole *VR* profile rather than a single *VR*, one is able to take advantage of additionally available information.

We have also shown that the Variance Ratio as described and used in the literature has a shortcoming when it comes to trending behaviour. As we have seen, this shortcoming can be easily overcome and the standard *VR* decisively improved.

Finally applying the modified *VR* method to the SP100 time series, we have shown that this portfolio is exhibiting structure one would not expect if the stock market followed a random walk. It is now the task of market practitioners to apply suitable methods to transform the detected structure into profit.

APPENDIX

A.1. SP100 movement in 1.000 days steps

The graphs below show the data clouds of the SP100 in a step-size of 1000 days. The time series in question are represented by an 'x' while the random cloud for comparison reasons is represented using spots. In addition, there are three randomly selected stocks highlighted as a square, a diamond and a star while the centre of the random cloud is a triangle (an arrow is pointing for better clarity). The ellipse marks the space where the random walk time series are located with a 90% probability. The dotted line roughly indicates the frontier between mean-reverting [left side] and trending behaviour [right side].

The SP100 time series move from the right bottom to the left top corner corresponding to a change from trending to mean-reverting behaviour. The highlighted three stocks seem to follow no specific pattern indicating that the structure is only visible when looking at the portfolio as a whole. Comparing these plots with those in figure 6, it is clear that the mean-reverting behaviour becomes more obvious the longer the length of the time series is.
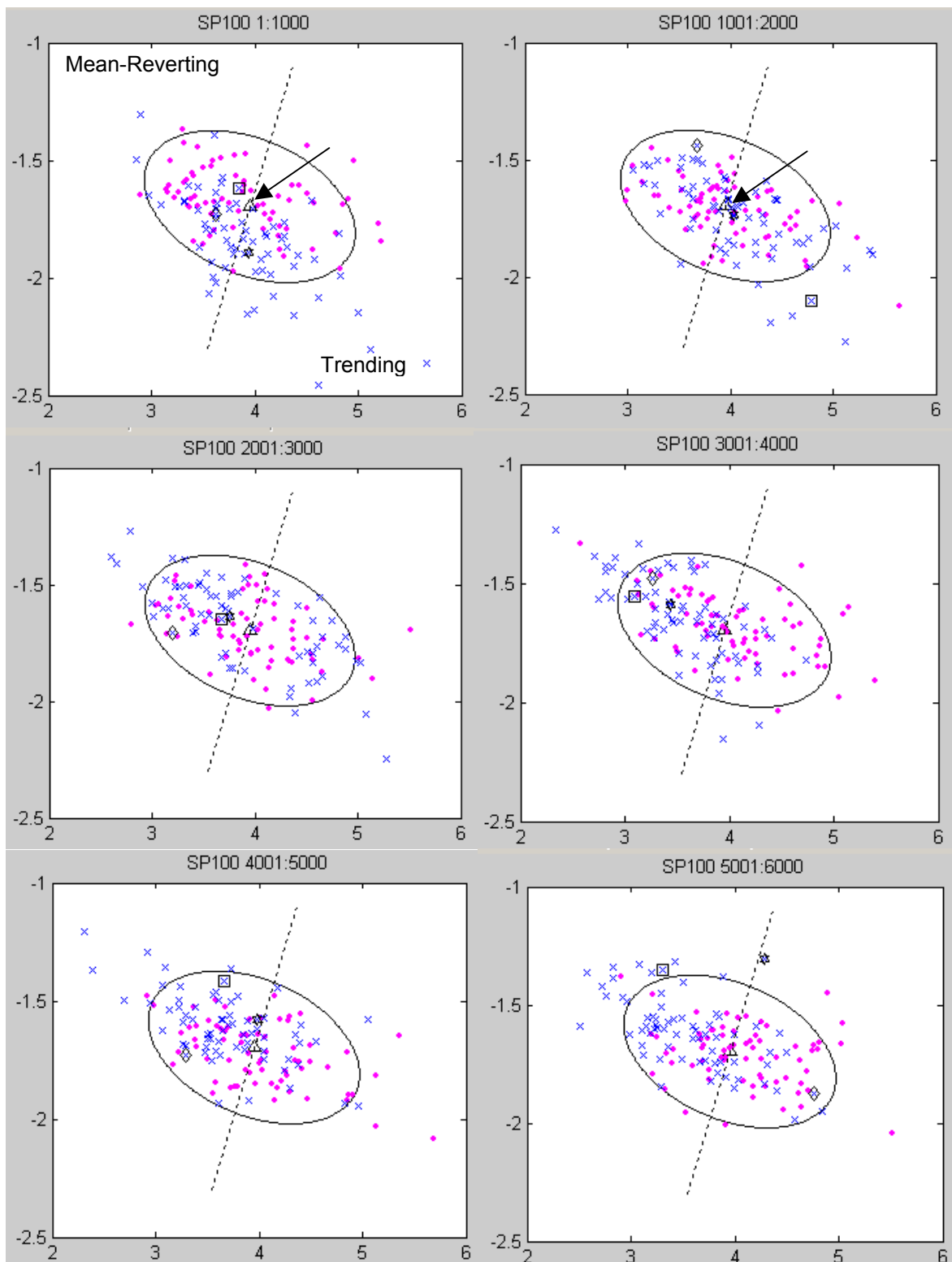
REFERENCES

Brock, W., J. Lakonishok and B. LeBaron (1992), 'Simple Technical Trading Rules and the Stochastic Properties of the Stock Returns', Journal of Finance, 47(5), 1731-1764.

Burgess, A.N.(1999), 'A Computational Methodology for Modelling the Dynamics of Statistical Arbitrage', PhD thesis, Decision Technology Centre, Department of Decision Science, University of London, London Business School.

Campbell, J.Y., A.W. Lo and A.C. MacKinlay (1997). *The Econometrics of Financial Markets.* Princeton University Press. Princeton, N.J.

Chow, K.V. and K.C. Denning (1993), 'A Simple Multiple Variance Ratio Test', Journal of Econometrics, 58, 385-401.

Fama, E.F. (1970), 'Efficient Capital Markets: A Review of Theory and Empirical Work', Journal of Finance, 25, 383-417.

Fama, E.F. (1991), 'Efficient Capital Markets II', Journal of Finance, 46, 1575-1617.

Fong, W.M., S.K. Koh and S. Ouliaris (1997), 'Joint Variance-Ratio Tests of the Martingale Hypothesis for Exchange Rates', Journal of Business and Economic Statistics, 15, 51-59.

Groenendijk, P.A., A. Lucas and C.G. de Vries (1998), 'A Hybrid Joint Moment Ratio Test for Financial Time Series', Research Memorandum 1998-47, Vrije Universiteit Amsterdam.

Jolliffe, I.T., (1986), *Principal Component Analysis*, Springer-Verlag, New York.

Levich, R.M. and L.R. Thomas (1993), 'The Significance of Technical Trading Rule Profits in the Foreign Exchange Market: A Bootstrap Approach', Journal of International Money and Finance, 12(5), 451-474.

Liu, C.Y. and J. He (1991), 'A Variance-Ratio Test of Random Walks in Foreign Exchange Rates', Journal of Finance, 46, 773-785.

Lo, A. and A.C. MacKinlay (1988), 'Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test', Review of Financial Studies, 1, 41-66.

Neely, C., P. Weller and R. Dittmar (1997), 'Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach', Journal of Financial and Quantitative Analysis, 32(4), 405-426.

Osler, C.L. and P.H.K. Chang (1995), 'Head and Shoulders: Not Just a Flaky Pattern', Federal Reserve Bank of New York Staff Report No 4.

Pagan, A.R. (1996). 'The Econometrics of Financial Markets', Journal of Empirical Finance, 3, 15-102.

Richardson, M. (1993), 'Temporary Components of Stock Prices: A Skeptic's View', Journal of Business and Economic Statistics, 11, 199-207.

Smith, G. and H.-J. Ryoo (2003), 'Variance Ratio Tests of the Random Walk Hypothesis for European Emerging Stock Markets', The European Journal of Finance, 9, 290-300.

Sullivan, R., A. Timmerman and H. White (1998), 'Data-Snooping, Technical Trading Rule Performance and the Bootstrap', Centre for Economic Policy Research Discussion Paper No. 1976.