Applied Financial Economics Letters, 2005, **1**, 189–197

Routledge
Taylor & Francis Group

Extending the variance ratio test to visualize structure in data: an application to the S&P 100 Index

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

[a] *Liverpool John Moores University*
[b] *CIBEF, Centre for International Banking, Economics and Finance, JMU, John Foster Building, 98 Mount Pleasant, Liverpool L3 5UZ*

The aim of this paper is to present a method able to graphically describe the amount of structure in a time series. In the following, 'structure' is defined as 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 is defined 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, trending is detected as well as mean-reverting structure over a time period of more than 10 years.

I. Introduction

The aim of this study is to present a method able to graphically describe the amount of structure in a time series. The following, defines 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 is defined as being trending respectively mean-reverting should be seen in relation to the characteristics of a random walk.

The motivation for the study 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 visualizing structure in data could be quite useful in the above-mentioned tasks.

Section II, gives a brief overview of the role that variance ratio tests play in the financial literature. Section III introduces and describes artificially created time series portfolios, which help one to assess the variance ratio method in a controlled environment. Section IV 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 V presents the idea of a modified variance ratio and shows its results when applied to the same data sets. Section VI 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 VII.

II. Literature Review

A lot of effort has been made to examine the efficiency of market price formation.[1] Most of the

* Corresponding author. E-mail: afealind@livjm.ac.uk
[1] Interesting surveys are provided by Fama (1970, 1991), Pagan (1996) and Campbell *et al.* (1997).

Applied Financial Economics Letters ISSN 1744–6546 print/ISSN 1744–6554 online

http://www.tandf.co.uk/journals
DOI: 10.1080/17446540500102315

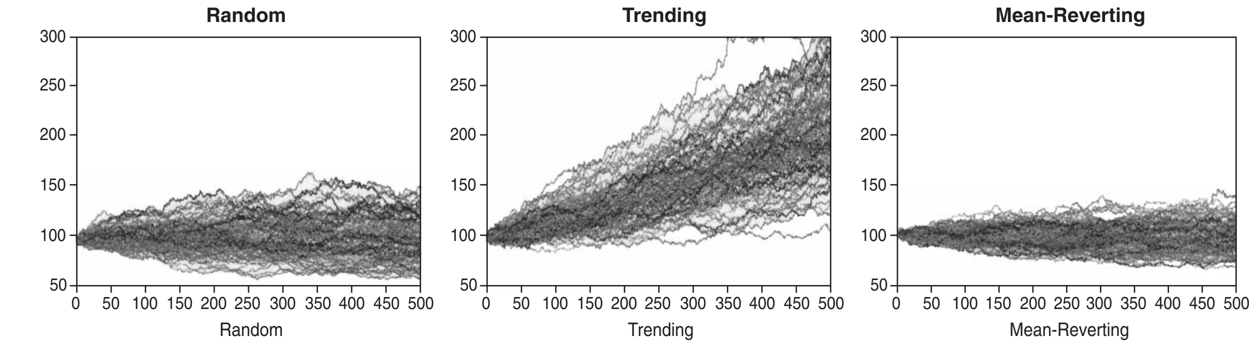

Fig. 1. Artificial time series

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 exchange 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 statistic 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 generalize the variance ratio equation to investigate what can be learned from using other moments in the ratio statistics besides 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.

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

III. Creating Artificial Data for a Controlled Environment

The following creates 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.

Figure 1 shows an overview of the different time series portfolios that are used. The x-axis represents time while the y-axis defines the levels of the series.[2]

[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.

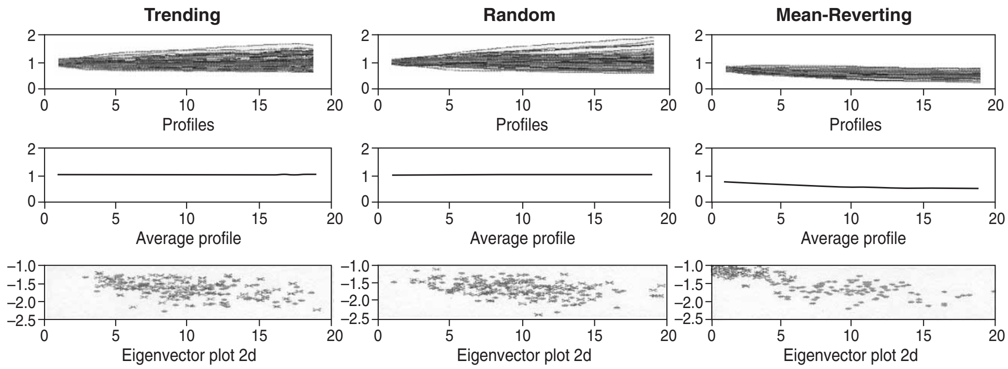

Fig. 2. Variance ratio profiles (for $\tau = 2 \ldots 20$) of the time series of the artificial portfolios

IV. The Variance Ratio

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]

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:

$$VAR\left(\Delta_{yt}^{\tau}\right) = \tau VAR(\Delta y_t) \quad (1)$$

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 $y_{t+1} - y_t$
- Δ_{yt}^{τ} 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) normalized by its single-period variance:

$$VR(\tau) = \frac{\sum_t \left(\Delta_{yt}^{\tau} - \overline{\Delta_y^{\tau}}\right)^2}{\tau \sum_t \left(\Delta y_t - \overline{\Delta y}\right)^2} \quad (2)$$

where

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

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

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:

$$VR(2) = 1 + \rho(1) \quad (3)$$

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 $VR(\rho)$ are specific combinations of the first k-1 autocorrelation coefficients of y_t, with linear declining weights (Campbell *et al.*, 1997).

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 the three artificial time series portfolios, the profiles using a long term variance ratio up to 20 days are shown in Fig. 2 where the x-axis specifies the τ of the VR and the y-axis shows the VR value.

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

[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).

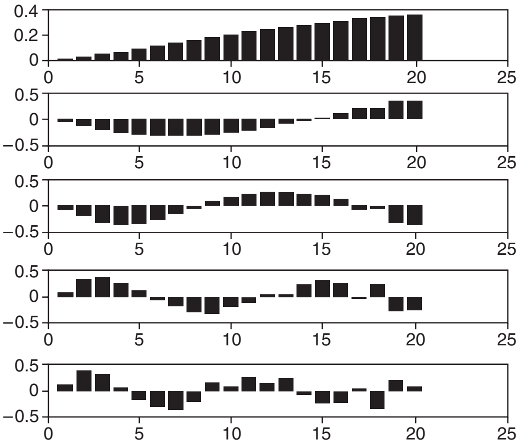

Fig. 3. The first five eigenvectors (archetypes) as a result of the application of PCA analysis to the co-variance matrix of random walk profiles

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 the present 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 the 'archetypes'.

The first five eigenvectors are shown in Fig. 3 (starting from top to bottom).

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 one to restrict oneself to the usage of the first two or three eigenvectors (respectively archetypes) while still being able to cover most of the structure. Using these few vectors enables one to plot one's results into either two or three 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) × (eigenvector)], and one is now able to characterize a profile by just two values: its similarity to the first and second eigenvectors.

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

Figure 4 analyses the artificial portfolios where Fig. 4a presents the trending, Fig. 4b the random and Fig. 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 boxes three and four. Here one finally get time series plotted as single points where the coordinates are dependent on the 'similarity' to the first two archetypes.

A single dot in the plot in box three represents a time series of one of the 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 boxes three and four are kept the same to simplify the comparability.

As can be seen on the graphs in Fig. 4, 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 as the random walk. This is a problem of the VR, which will be studied closer in the next section.

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.

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

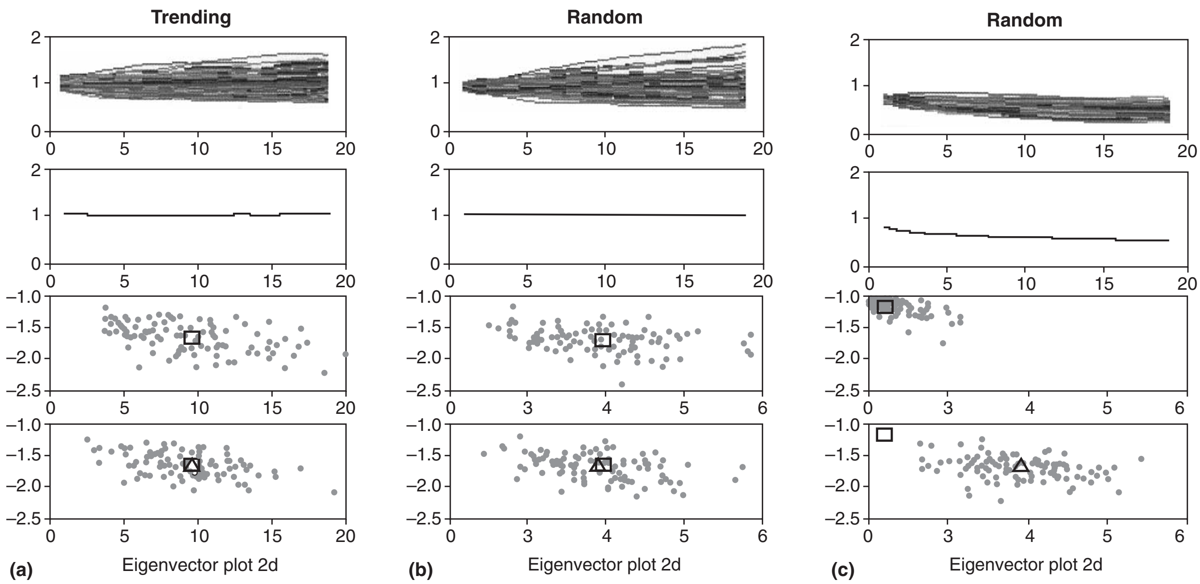

Fig. 4. The artificial time series are mapped as spots into two dimensional space and compared to a random walk series

To find out why, two trending time series are defined. One is created by sampling from a normal distribution N(0,1) where the returns are rearranged in a way that produces 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, one has 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 one has to normalize 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 Fig. 4a and b look the same.

V. The Modified Variance Ratio

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:

$$VR^{*}(\tau) = \frac{\sum_t \left(\Delta_{yt}^{\tau}\right)^2}{\tau \sum_t \left(\Delta y_t\right)^2} \tag{4}$$

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.

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, second, the mean-reverting series are plotted to the left side, while trending series are more to the right side.

VI. Applying the Modified Variance Ratio to the S&P100 Index

It has been 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

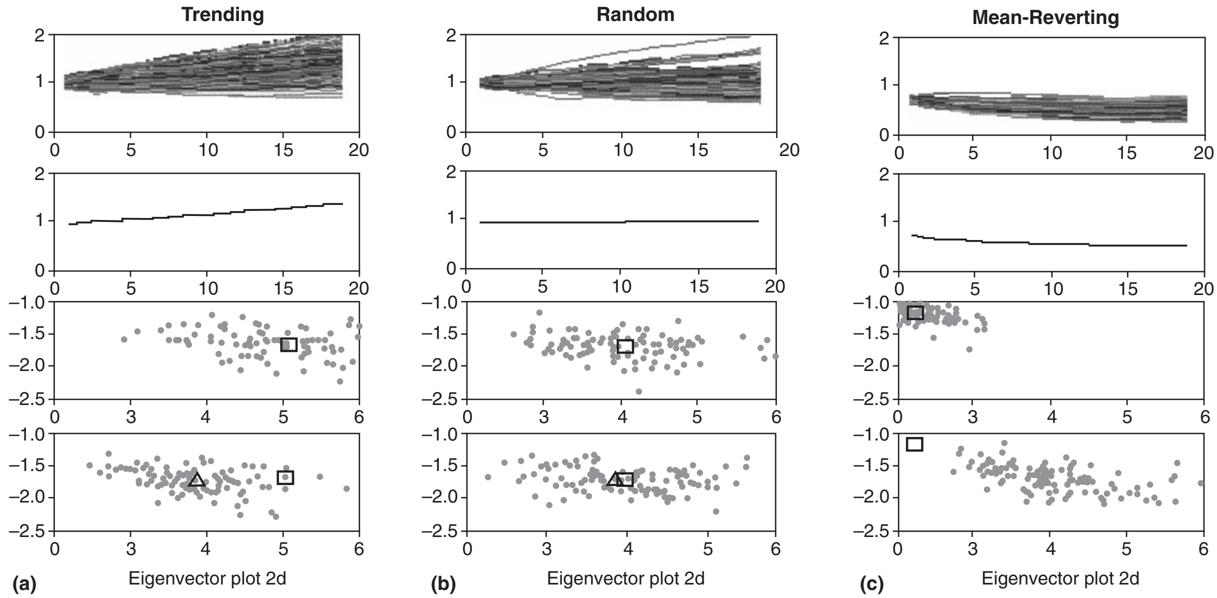

Fig. 5. Mapping the artificial time series into two dimensional space using the modified VR Equation 4 to calculate the profiles

Table 1. The EUR/USD dataset

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

but also to use the whole profile rather than a single pre-defined VR. The methodology is applied 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 and is shown in Table 1.

This long time period reduces the time series available from 100 to 69. These remaining companies are shown in Table 2.

The profiles, i.e. the eigenvector plot for this portfolio, are presented in Fig. 6. 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, each time series has been smoothed by using a rolling window of 50 days length to normalize 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 the 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 Fig. 6a).

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 in Fig. 7. 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.

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.

Table 2. Components of training dataset

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

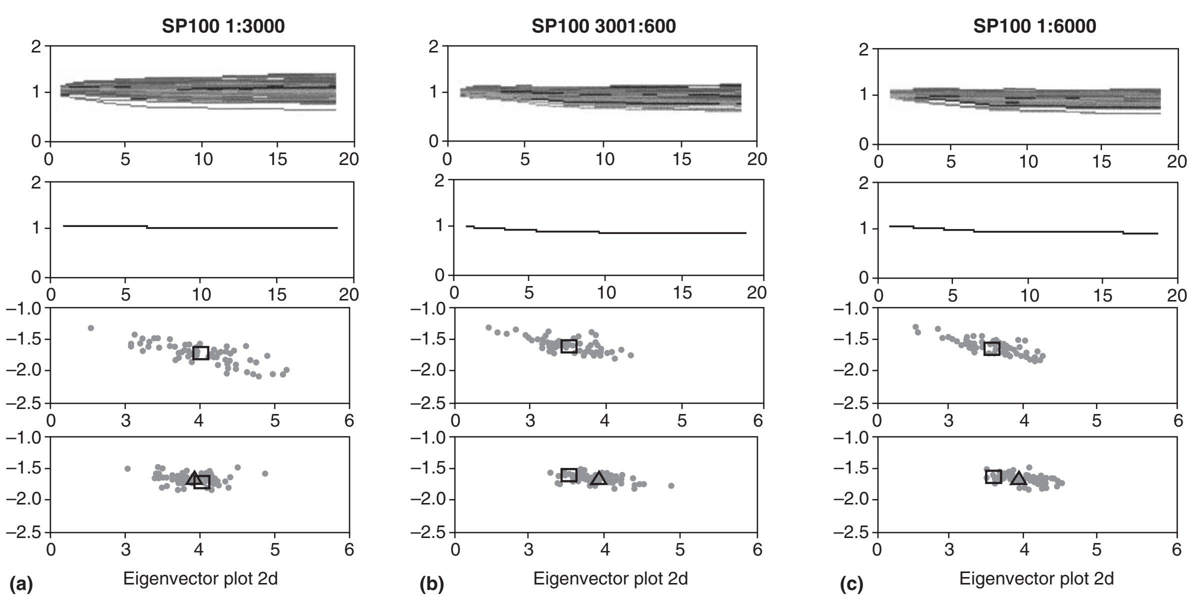

Fig. 6. Plotting the S&P constituents into two dimensional space and comparing them with a random walk portfolio

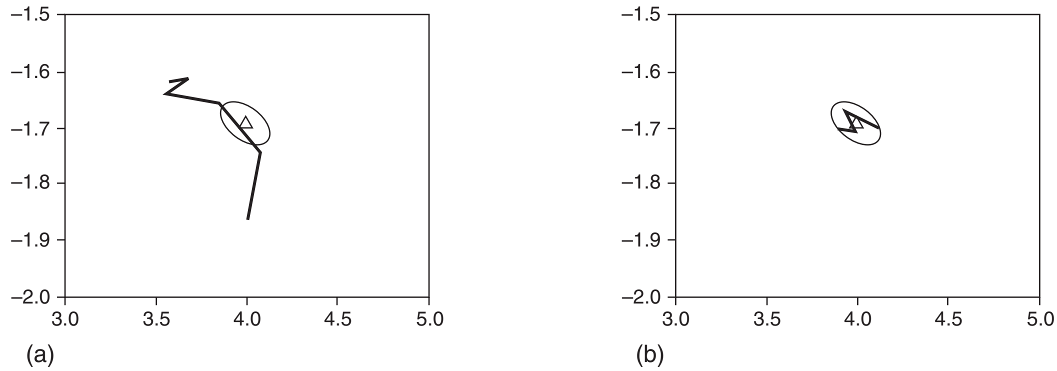

Fig. 7. Tracking the movements of the centre of the S&P constituents cloud (a) compared to the movements of a random walk portfolio (b)

VII. Concluding Remarks

It has been 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.

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

Finally applying the modified VR method to the SP100 time series, it has been 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.

References

Brock, W., Lakonishok, J. and LeBaron, B. (1992) Simple technical trading rules and the stochastic properties of the stock returns, *Journal of Finance*, **47**(5), 1731–64.

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

Chow, K. V. and Denning, K. C. (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**, 1575–617.

Fong, W. M., Koh, S. K. and Ouliaris, S. (1997) Joint variance-ratio tests of the Martingale hypothesis for exchange rates, *Journal of Business and Economic Statistics*, **15**, 51–9.

Groenendijk, P. A., Lucas, A. and de Vries, C. G. (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 Thomas, L. R. (1993) The significance of technical trading rule profits in the foreign exchange market: a bootstrap approach, *Journal of International Money and Finance*, **12**(5), 451–74.

Liu, C. Y. and He, J. (1991) A variance-ratio test of random walks in foreign exchange rates, *Journal of Finance*, **46**, 773–85.

Lo, A. and MacKinlay, A. C. (1988) Stock market prices do not follow random walks: evidence from a simple specification test, *Review of Financial Studies*, **1**, 41–66.

Neely, C., Weller, P. and Dittmar, R. (1997) Is technical analysis in the foreign exchange market profitable? A genetic programming approach, *Journal of Financial and Quantitative Analysis*, **32**(4), 405–26.

Osler, C. L. and Chang, P. H. K. (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 Ryoo, H.-J. (2003) Variance ratio tests of the random walk hypothesis for European emerging stock markets, *The European Journal of Finance*, **9**, 290–300.

Sullivan, R., Timmerman, A. and White, H. (1998) Data-snooping, technical trading rule performance and the bootstrap, Centre for Economic Policy Research Discussion Paper No. 1976.

Appendix A1. SP100 Movement in 1000 Days Steps

Figure A1 shows 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 Fig. 6, it is clear that the mean-reverting behaviour becomes more obvious the longer the length of the time series is.

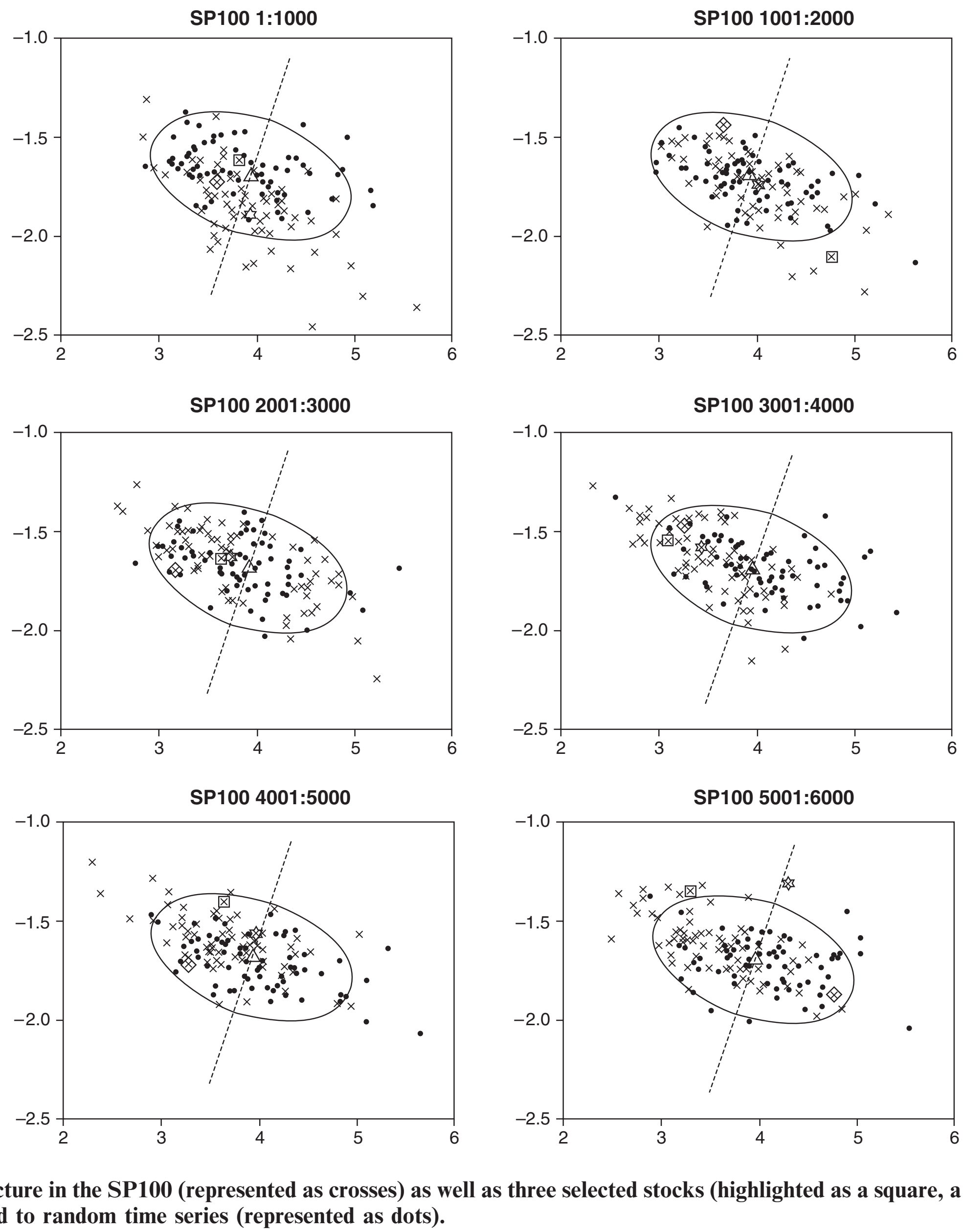

Fig. A1. Structure in the SP100 (represented as crosses) as well as three selected stocks (highlighted as a square, a diamond and a star) compared to random time series (represented as dots).