Clustering financial time series with variance ratio statistics[*]

João A. Bastos and Jorge Caiado[†]

CEMAPRE, ISEG, Technical University of Lisbon, 1200-781 Lisboa, Portugal

Forthcoming: *Quantitative Finance*

Abstract

This study introduces a new distance measure for clustering financial time series based on variance ratio test statistics. The proposed metric attempts to assess the level of interdependence of time series from the point of view of return predictability. Simulation results show that this metric aggregates better time series according to their serial dependence structure than a metric based on the sample autocorrelations. An empirical application of this approach to international stock market returns is presented. The results suggest that this metric discriminates reasonably well stock markets according to size and level of development. Furthermore, despite the substantial evolution of individual variance ratio statistics, the clustering pattern remains fairly stable across different time periods.

1 Introduction

Clustering of time series has become an important tool in many scientific domains, such as finance and economics, engineering and life sciences. The procedure for clustering time series typically involves the construction of a convenient similarity measure between the series. A number of approaches for clustering time series data are available in the literature, such as autoregressive expansion-based distances (Piccolo, 1990; Maharaj, 1996, 1999,

[*]An earlier version of this paper was circulated under the title "Clustering global equity markets with variance ratio tests".

[†]Corresponding author. Email: jcaiado@iseg.utl.pt

2000), autocorrelation-based distances (Galeano and Peña, 2000; Caiado et al., 2006), fitted residuals-based distances (Tong and Dabas, 1990), cross-correlation coefficient distances (Bohte et al., 1980), periodogram-based distances (Maharaj, 2002; Caiado et al., 2006, 2009), spectral coherence-based dissimilarities (Maharaj and D'Urso, 2010), dynamic time warping distances (Berndt and Clifford, 1996; Wang and Gasser, 1997), Markov-operator distances (Gregorio and Iacus, 2008), short time series distances (Möller-Levet et al., 2003) and cepstral coefficient-based distances (Kalpakis et al., 2001; Savvides et al., 2008).

Cluster analysis of time series is particularly important in finance, since practitioners are interested in identifying similarities in financial assets for investment and risk management purposes. This has motivated financial researchers to develop multivariate statistical methods to identify similar structural patterns in asset prices. For instance, Mantegna (1999) and Bonanno et al. (2001) used a function of the Pearson correlation coefficient as a measure of similarity between pairs of stock returns. In order to take into account the information about the volatility structure of time series, Caiado and Crato (2010) introduced a Mahalanobis-like distance between the dynamic features of two return series and employed a clustering procedure to investigate similarities among stocks of the DJIA index.

In this paper, we introduce a new distance measure for clustering time series with similar stochastic dependence structure. The proposed metric is based on the distance between variance ratio statistics computed for individual series. Variance ratios tests are popular tests of the hypothesis that a time series follows a random walk or a martingale difference sequence, i.e., that its returns are uncorrelated at all leads and lags. Lo and MacKinlay (1988, 1989) provided the asymptotic sampling theory for both homoscedastic and heteroscedastic random walks, and showed that these tests are more powerful than traditional tests, such as serial correlation and unit root tests, against several alternative processes. Wright (2000) proposed non-parametric variance ratio tests based on ranks and signs. Unlike conventional variance ratio tests, rank- and sign-based tests are exact, with sampling distributions that do not rely on asymptotic approximations. Wright (2000) showed that rank- and sign-based tests improve substantially the power of variance ratio tests with little size distortions. In recent years, many innovations and refinements of the variance ratio methodology have been proposed in the literature. An extensive survey of these developments is provided by Charles and Darné (2009).

To some extent, the proposed metric assesses the level of interdependence of time-series from the point of view of return predictability. Therefore, a natural empirical application for this metric is provided by international stock markets. In fact, the level of predictability of global markets is of great importance for investors seeking the reduction of idiosyncratic risk through international portfolio diversification, and an abundant research has been devoted to examining whether the prices of securities conform to a random walk behavior. A comprehensive review of these empirical studies is provided in the recent survey by Lim and Brooks (2011). In particular, several studies employed variance ratios to test the random walk hypothesis in stock markets (see, e.g., Hoque et al., 2007; Kim and Shamsuddin, 2008; Smith, 2009).

In this study, we analyze daily returns of free float-adjusted market capitalization equity indices from 46 different countries, covering the period from 1995 to 2009. Two types of multivariate interdependence techniques are considered for analyzing the clustering patterns of these markets. First, we employ multidimensional scaling maps, which can be used to identify similarities between features of different return series, and to construct distances in a multidimensional space. Then, we perform cluster analysis, which is particularly suited for defining groups of equity markets with maximal structure dependence within the groups while also having minimum structure dependence between the groups.

The remainder of this paper is organized as follows. Section 2 presents a brief review of the variance ratio statistics that are used as inputs for our metric. Section 3 introduces the clustering procedure and evaluates its properties on simulated data. Section 4 shows the results of an empirical application to international stock markets. Finally, Section 5 presents some concluding remarks.

2 Variance ratio tests

In this section, we describe three alternative variance ratio tests of the random walk hypothesis. If a return series conforms to a random walk behavior then it should be uncorrelated at all leads and lags. Let p_t, $t = 0, 1, ..., T$ denote a time-series of asset prices and y_t denote the continuously compounded return at time t, $y_t = \log(p_t/p_{t-1})$, $t = 1, ..., T$. Given the time series of asset returns $y_t = \mu + \epsilon_t$, where μ is a drift parameter, we want to test the null hypothesis that: i) ϵ_t are identical and independently distributed (iid), or ii) ϵ_t are independent and conditional heteroscedastic, that is, the return series forms a martingale difference sequence.

2.1 Conventional variance ratio tests

Lo and MacKinlay (1988) variance ratio tests are based on the property that, if returns are i.i.d., the variance of the k-period return is k times the variance of the one-period return. Therefore, if a return series is a random walk the ratio of $1/k$ times the variance of $\log(p_t/p_{t-k})$ to the variance of $\log(p_t/p_{t-1})$ should be close to 1. This variance ratio is given by

$$\mathrm{VR}(k) = \frac{\frac{1}{Tk}\sum_{t=k}^{T}\left(y_t + y_{t-1} + ... + y_{t-k+1} - k\hat{\mu}\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{\mu}\right)^2}, \tag{1}$$

where $\hat{\mu} = T^{-1}\sum_{t=1}^{T} y_t$. Lo and MacKinlay (1988) showed that, if the returns are i.i.d. then the test statistic

$$M_1(k) = (\mathrm{VR}(k) - 1)\,\phi(k)^{-1/2}, \tag{2}$$

where

$$\phi(k) = \frac{2(2k-1)(k-1)}{3kT}, \tag{3}$$

follows the standard normal distribution asymptotically, under the null hypothesis that $\mathrm{VR}(k) = 1$. This null asymptotic distribution does not hold if the returns are subject to conditional heteroscedasticity. Therefore, Lo and MacKinlay (1988) proposed an alternative test statistic which is robust against the presence of conditional heteroscedasticity, given by

$$M_2(k) = (\mathrm{VR}(k) - 1)\left[\sum_{j=1}^{k-1}\left[\frac{2(k-j)}{k}\right]^2 \delta_j\right]^{-1/2}, \tag{4}$$

where

$$\delta_j = \frac{\sum_{t=j+1}^{T}(y_t - \hat{\mu})^2(y_{t-j} - \hat{\mu})^2}{\left[\sum_{t=1}^{T}(y_t - \hat{\mu})^2\right]^2}. \tag{5}$$

The test statistic $M_2(k)$ also follows the standard normal distribution asymptotically under the null hypothesis that $\mathrm{VR}(k) = 1$ and the conventional critical values for the standard normal distribution hold for both tests.

2.2 Rank-based variance ratio tests

The finite-sample null distribution of Lo and MacKinlay (1988) tests can be rather asymmetric and nonnormal, exhibiting bias and positive skewness. To overcome these problems, Wright (2000) proposed non-parametric tests based on ranks and signs. These tests have exact sampling distributions and do not recur to any asymptotic approximations. Denote by $r(y_t)$ the rank of y_t among $y_1, ..., y_T$. Under the null hypothesis that ϵ_t is i.i.d., the integer sequence $r(y_t)$, $t = 1, ..., T$, is a random permutation of the integers from 1 to T, where each permutation has equal probability. Wright (2000) suggests two alternative standardizations of the ranks,

$$r_{1t} = \frac{r(y_t) - \frac{T+1}{2}}{\sqrt{\frac{T^2-1}{12}}} \tag{6}$$

and

$$r_{2t} = \Phi^{-1}\left(\frac{r(y_t)}{T+1}\right), \tag{7}$$

where Φ is the standard normal cumulative distribution function. The proposed test statistics are given by

$$R_1(k) = \left[\frac{\frac{1}{Tk}\sum_{t=k}^{T}\left(r_{1t} + r_{1t-1} + ... + r_{1t-k+1}\right)^2}{\frac{1}{T}\sum_{t=1}^{T} r_{1t}^2} - 1\right] \times \phi(k)^{-1/2}, \tag{8}$$

and

$$R_2(k) = \left[\frac{\frac{1}{Tk}\sum_{t=k}^{T}\left(r_{2t} + r_{2t-1} + ... + r_{2t-k+1}\right)^2}{\frac{1}{T}\sum_{t=1}^{T} r_{2t}^2} - 1\right] \times \phi(k)^{-1/2}. \tag{9}$$

The exact sampling distribution of $R_1(k)$ and $R_2(k)$ may be derived from simulation to any arbitrary degree of accuracy. Note that in the presence of conditional heteroscedasticity $r(y_t)$, $t = 1, ..., T$, no longer corresponds to a random permutation of the set $1, ..., T$ with equal probability and $R_1(k)$ and $R_2(k)$ are not exact. However, through Monte Carlo simulations Wright (2000) showed that these tests do not exhibit serious size distortions under conditional heteroscedasticity.

2.3 Sign-based variance ratio tests

Wright (2000) also suggests a sign-based variance ratio test. Let

$$s_t = \begin{cases} 1 & \text{if } y_t > 0 \\ -1 & \text{otherwise} \end{cases} \tag{10}$$

If $\mu = 0$, the series s_t is i.i.d. with mean 0 and variance 1. Also, each s_t is equal to 1 with probability $\frac{1}{2}$ and equal to -1 otherwise. The sign-based test statistic is given by

$$S_1(k) = \left[\frac{\frac{1}{Tk} \sum_{t=k}^{T} \left(s_t + s_{t-1} + ... + s_{t-k+1} \right)^2}{\frac{1}{T} \sum_{t=1}^{T} s_t^2} - 1 \right] \times \phi(k)^{-1/2}. \tag{11}$$

Again, the exact sampling distribution of $S_1(k)$ may be obtained by simulation. This test is exact even in the presence of conditional heteroscedasticity. Because the assumption that $\mu = 0$ is restrictive, Wright (2000) suggested an alternative test in which this condition is relaxed but his Monte Carlo simulations showed that the power of this test did not compare well with $S_1(k)$.

3 Cluster analysis with variance ratio statistics

3.1 Variance ratio-based metric

A fundamental task in cluster analysis is to obtain a relevant measure of similarity between each pair of time series. Here, we propose the Euclidean distance between vectors of the variance ratio statistics M_1, M_2, R_1, R_2 and S_1, introduced in Section 2. Furthermore, these vectors include variance ratios evaluated at several lags k in order to capture the serial dependence of the returns. To eliminate any bias due to scale differences across variables, the variance ratios are standardized before computing the distances. Denoting by $v'_x = [\mathrm{VR}_{1x}, \mathrm{VR}_{2x}, ..., \mathrm{VR}_{px}]$ and $v'_y = [\mathrm{VR}_{1y}, \mathrm{VR}_{2y}, ..., \mathrm{VR}_{py}]$ the p-dimensional vectors of standardized variance ratios for time series x and y, respectively, the distance measure between these vectors is

$$d_{\mathrm{VR}}(x, y) = \sqrt{\sum_{j=1}^{p} (\mathrm{VR}_{jx} - \mathrm{VR}_{jy})^2} \tag{12}$$

Let n denote the number of time series under consideration. We compute dissimilarities between every pair of series in the data set. The result of this computation is an Euclidean distance matrix D with $n(n-1)/2$ different pairs of time series.

3.2 Multidimensional scaling

Multidimensional scaling (MDS) is a multivariate statistical method that uses the information about the similarities (or dissimilarities) between objects to construct a configuration of n points in low-dimensional space (see, for instance, Johnson and Wichern, 2007). Let D be the observed $n \times n$ matrix of Euclidean distances. By multidimensional scaling, the matrix D yields a $n \times d$ configuration matrix T. The rows of T are the coordinates of the n points in a d-dimensional representation of the observed dissimilarities ($d < n$). The d-dimensional representation that best approximates the observed dissimilarity matrix is given by the d eigenvectors of TT' corresponding to the d largest eigenvalues.

3.3 Hierarchical cluster analysis

Cluster analysis attempts to determine groups (or clusters) of objects in a multivariate data set. The most commonly used clustering algorithm is based on the hierarchical classification of the objects. This linkage algorithm is concerned with the partition of a set of objects into groups or clusters, in such a way that objects in the same group are similar to one another and objects in different clusters are as distinct as possible. We begin with each object being considered as a separate cluster (n clusters). In the second stage, the closest two groups are linked to form $n - 1$ clusters. The process continues until the last stage, in which all the objects are in the same cluster (for further discussion, see Johnson and Wichern, 2007).

The dendrogram (also called "cluster tree") is a graphical representation of the results of the hierarchical cluster analysis. The dendrogram shows how clusters are formed at each stage of the procedure. At the bottom of the dendrogram, each object is considered its own cluster. The objects continue to combine upwards. At the top, all objects are grouped into a single cluster. In hierarchical clustering, partitions are obtained by cutting off the dendrogram at an arbitrary point. The choice of the appropriate number of clusters in the dendrogram is sometimes subjective and depends on the expert judgment of the researchers. A formal method for finding the appropriate partition in the data set are the Duda-Hart Je(2)/Je(1) indices (Duda and Hart, 1973). These indices (also called "stopping rules") are computed for each cluster solution in a hierarchical cluster analysis. Larger values of the indices indicate more distinct clustering (for more details, see Everitt et al., 2001).

3.4 Simulation results

In order to understand the properties of the variance-ratio based distance $d_{\mathrm{VR}}(x, y)$ in clustering time series, we simulated 10 random series from each of the following six processes (see Wright, 2000):

Model (a): $y_t = \varepsilon_t$, where ε_t is iid N(0,1).

Model (b): $y_t = \exp(h_t/2)\varepsilon_t$, where $h_t = 0.95h_{t-1} + \xi_t$, with ξ_t iid N(0,0.1) and independent of ε_t.

Model (c): $y_t = 0.1y_{t-1} + \exp(h_t/2)\varepsilon_t$;

Model (d): $(1 - L)^d y_t = \exp(h_t/2)\varepsilon_t$, where $(1 - L)^d$ is the fractional differencing operator and $d = 0.1$;

Model (e): $y_t = (1 - 0.5L)^{-1}v_t + \exp(h_t/2)\varepsilon_t$, where v_t is iid N(0,0.1), independent of ε_t;

Model (f): $y_t = (1 - L)^{-d}v_t + \exp(h_t/2)\varepsilon_t$, where $(1 - L)^d$ is the fractional differencing operator and $d = 0.3$.

For each series we calculated the variance ratios M_1, M_2, R_1, R_2 and S_1 evaluated at four distinct values of k: 2, 5, 10, 20. Then the Euclidean distance $d_{\mathrm{VR}}(x, y)$ between all pairs of 20-dimensional vectors was computed. We also considered three different series lengths: $N = 500$, 1000 and 2000. We explore the existence of possible clusters among the six models by the complete linkage dendrogram associated with the variance ratio distances. The dendrograms from which clusters can be identified are shown in Figure 1. For large values of N, the method tends to group the series according to serial dependence structure. For example, one cluster contains almost all series generated from processes (c) and (d). These two processes have both serial dependence and conditional heteroscedasticity. On the other hand, the series generated from processes with no serial dependence (Models (a) and (b)), the series generated from the sum of a Gaussian AR(1) process and conditional heteroscedastic noise (Model (e)), and the series generated from the sum of Gaussian fractionally integrated process with conditional heteroscedastic noise (Model (f)) are, to a large degree, randomly distributed across multiple clusters.

For comparison, we also consider a well-known discrepancy statistic based on the esti-

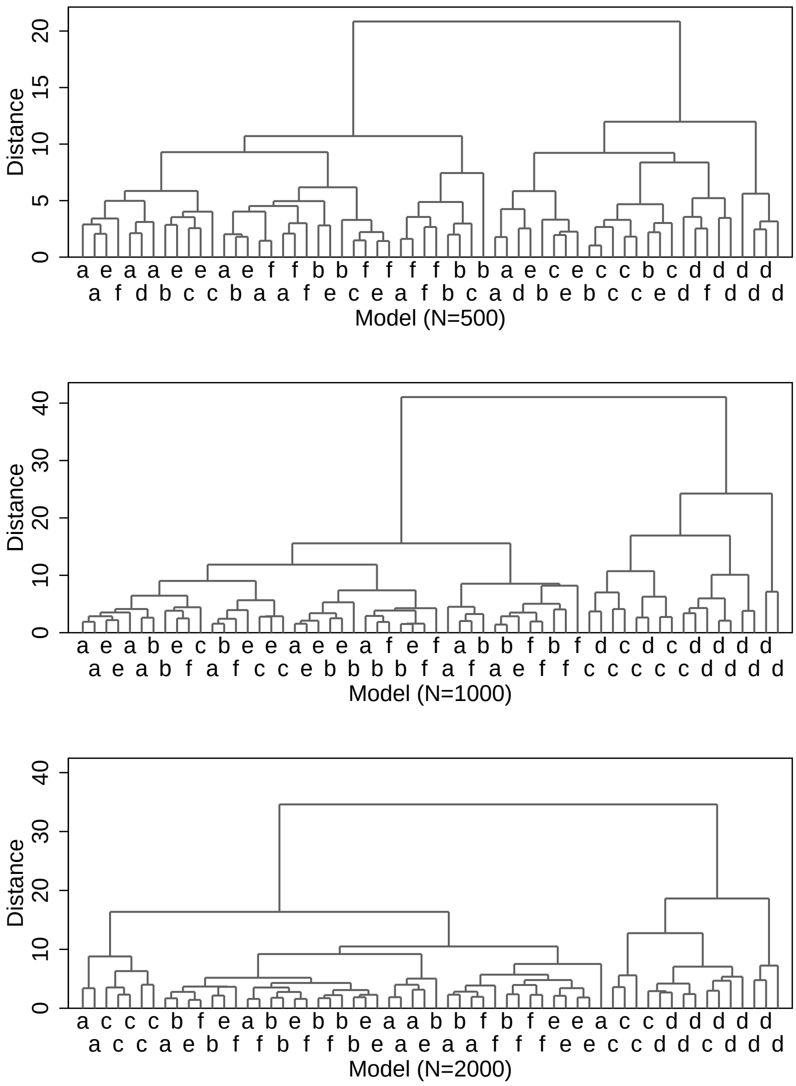

Figure 1: Complete linkage dendrogram for variance-ratio based distances between the generated series.

mated autocorrelations (Galeano and Peña, 2000):

$$d_{\mathrm{ACF}}(x,y) = \sqrt{\sum_{l=1}^{L} \left(\widehat{\rho}_{x,l} - \widehat{\rho}_{y,l}\right)^2}, \tag{13}$$

where $\widehat{\rho}_{x,l}$ and $\widehat{\rho}_{y,l}$ are the sample autocorrelation functions of time series x and y, and L is the number of autocorrelation lags (in our simulation study, we set $L = N/10$, as recommended in Caiado et al. (2006)). Figure 2 shows the complete linkage dendrogram based on metric (13). From the results given by these simulations, it can be seen that the autocorrelation method performs poorly in detecting serial dependence. In fact, this metric cannot distinguish models that have serial dependence and some conditional heteroscedasticity from those that are iid or mds.

In order to better assess the methods, we have explored other hierarchical (single linkage, average linkage and Ward's linkage) and non-hierarchical (k-means) clustering procedures. Irrespectively of the clustering procedure, the variance ratio based metric provides better cluster solutions than the ACF based method. We have also obtained multidimensional scaling solutions for both variance ratio and ACF distances between generated series. Results were similar to the ones obtained by the hierarchical and non-hierarchical clustering procedures and provide the same recommendations for the discrepancy statistic choice.

4 An empirical application

4.1 Data

The data employed in this analysis consists of free float-adjusted market capitalization equity indices constructed and maintained by Morgan Stanley Capital International (MSCI). In order to avoid effects due to exchange rates, all indices are specified in local currency. The construction and maintenance of the MSCI index family follows a consistent methodology. Securities included in the indices are subject to minimum requirements in terms of market capitalization, free-float, liquidity, availability to foreign investors and length of trading.

The database includes the following 23 developed markets: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Hong Kong, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, the United Kingdom and the United States. A market is classified as developed if i) the country GNI per capita is 25% above the World Bank high income threshold for 3 consecutive

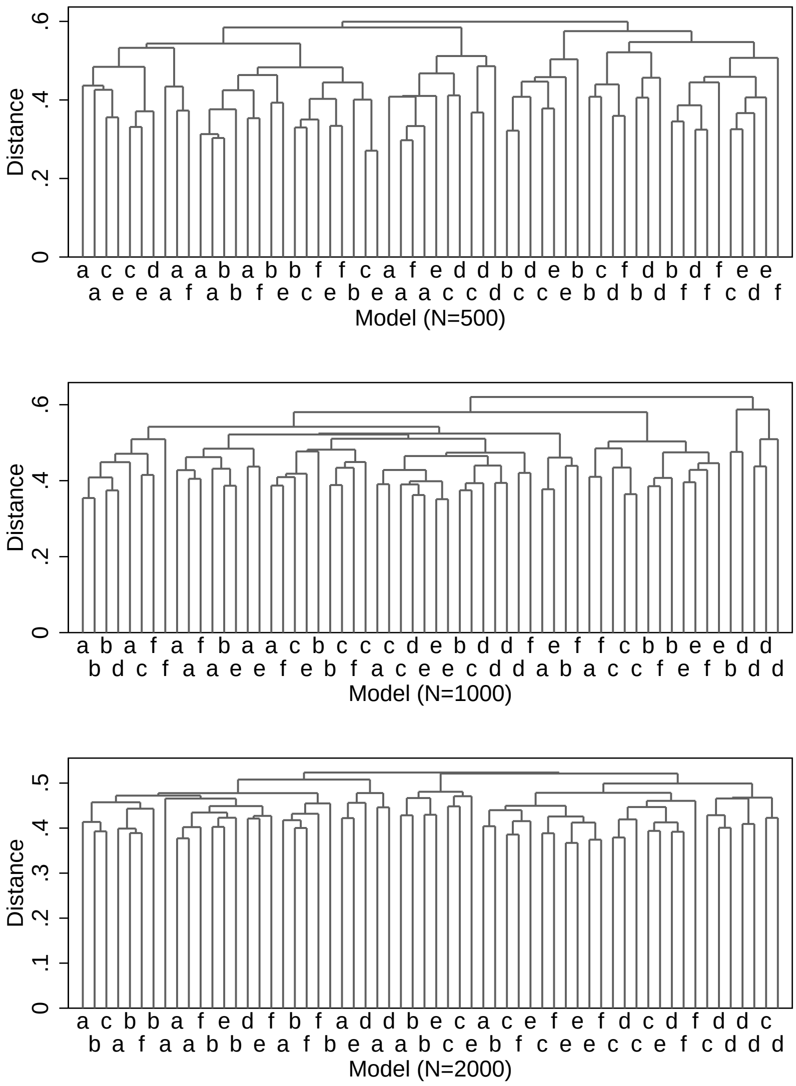

Figure 2: Complete linkage dendrogram for autocorrelation-based based distances between the generated series.

years, ii) there is a minimum number of companies satisfying minimum size and liquidity requirements, and iii) there is a very high openness to foreign ownership, ease of capital inflows/outflows, efficiency of the operational framework and stability of institutional framework. The database also includes the following 23 emerging markets: Argentina, Brazil, Chile, China, Czech Republic, Colombia, Egypt, Hungary, India, Indonesia, Israel, Korea, Malaysia, Mexico, Morocco, Peru, Philippines, Poland, Russia, South Africa, Taiwan, Thailand and Turkey. To be included in the emerging category, a market must satisfy size and liquidity requirements, and market accessibility criteria that are less tight than those for their counterparts in developed economies.[1]

Although for some developed markets the database includes observations that date back to year 1979, only observations from year 1995 onwards are available for the complete set of developed and emerging markets. The data employed in the analysis consists of daily index prices from 1995:01 to 2009:12, corresponding to 3,914 observations.[2] In the event of days where there is a market holiday, the index construction methodology simply carries forward the index value from the previous business day. In some countries the number of market holidays can be quite large (for example, the data for Egypt includes 914 market holidays). Therefore, repeated observations were removed from the data in order to remove potential biases associated to nontrading days.

Summary statistics (size, annualized mean, annualized standard deviation, skewness and kurtosis coefficients) of daily percentage rates of return, $\log(p_t/p_{t-1}) \times 100$, where p_t is the index price at time t, for developed and emerging markets are reported in Tables 1 and 2, respectively. The mean return and the standard deviation of the returns for emerging markets (10.07% and 29.33%) are higher than those for developed markets (4.29% and 22.95%), reflecting the risk/return trade-off suggested by finance theory. Several markets (i.e., Ireland, Japan, Norway, China, Philippines, Taiwan and Thailand) exhibit negative mean returns in this period. Most developed markets exhibit negative skewness coefficients, indicating that return distributions in these markets typically have longer negative than positive tails. On the other hand, almost half of the emerging markets exhibit positive skewness. The returns series for the developed markets of Austria, Belgium, Canada, Hong-Kong, Ireland, Norway, New Zealand, Portugal and United States, and for the emerging markets of Brazil, Chile, Colombia, Czech Republic, Hungary, Indonesia, Malaysia, Philippines, Russia and Thailand are highly leptokurtic, which means that extreme events occur with increasing frequency.

[1]For details see `www.mscibarra.com`

[2]Note that this sample covers the period posterior to the movement towards financial liberalization experienced by many emerging economies in the late 80's and early 90's (see Kim and Singal, 2000).

Market	size	mean (%)	stdev (%)	skew	kurt
Australia	3797	6.02	16.62	-0.418	9.071
Austria	3713	1.78	22.85	-0.349	13.344
Belgium	3803	1.75	21.44	-0.588	14.154
Canada	3776	8.12	20.08	-0.636	11.563
Denmark	3755	8.89	20.25	-0.332	9.028
Finland	3758	7.71	37.43	-0.360	8.969
France	3804	5.52	22.93	-0.063	7.706
Germany	3795	4.62	24.56	-0.101	7.383
Greece	3742	4.29	27.70	-0.118	7.074
Hong-Kong	3703	3.89	27.02	0.026	11.434
Ireland	3777	-3.18	25.55	-0.722	15.672
Italy	3796	2.85	22.68	-0.062	7.916
Japan	3686	-3.25	22.44	-0.133	8.445
Netherlands	3819	4.10	23.21	-0.182	7.998
Norway	3767	-0.92	18.14	-0.632	18.228
New Zealand	3763	5.83	25.02	-0.564	10.054
Portugal	3770	4.09	17.76	-0.278	11.423
Singapore	3765	1.83	22.80	0.030	8.460
Spain	3774	10.15	23.29	-0.132	7.834
Sweden	3762	8,67	26.66	0.091	6.398
Switzerland	3769	6.25	19.73	-0.102	8.561
United Kingdom	3792	3.60	19.31	-0.174	9.399
United States	3778	6.00	20.34	-0.214	11.125
Average	3768	4.29	22.95	-0.261	10.054

Table 1: Summary statistics of daily percentage rates of return for developed markets: number of observations (size), annualized mean (mean), annualized standard deviation (stdev), skewness (skew) and kurtosis (kurt).

Market	size	mean (%)	stdev (%)	skew	kurt
Argentina	3721	12.58	37.11	-0.066	9.488
Brazil	3711	15.44	33.56	0.333	13.417
Chile	3741	6.62	18.07	0.331	13.081
China	3834	-0.56	34.22	0.035	8.074
Colombia	3637	16.94	22.98	0.178	14.630
Czech Republic	3741	8.52	25.18	-0.359	12.167
Egypt	3000	16.96	26.19	-0.251	7.940
Hungary	3744	16.76	31.95	-0.379	11.200
India	3683	10.42	27.65	-0.139	8.311
Indonesia	3659	11.65	33.97	-0.130	11.277
Israel	3890	8.70	23.62	-0.354	7.843
Korea	3703	6.50	34.07	0.020	6.590
Malaysia	3697	2.11	25.35	0.763	44.721
Mexico	3771	16.44	25.97	0.099	7.844
Morocco	3668	8.69	13.29	-0.063	9.031
Peru	3744	14.97	29.08	-0.136	9.618
Philippines	3706	-1.24	25.95	0.350	12.333
Poland	3755	6.60	29.39	-0.115	5.159
Russia	3841	13.40	52.17	-0.367	12.633
South Africa	3746	9.06	21.42	-0.413	7.811
Taiwan	3690	-0.11	26.92	-0.024	5.030
Thailand	3675	-4.10	37.75	0.661	13.321
Turkey	3731	35.36	43.72	0.034	7.532
Average	3699	10.07	29.33	0.000	11.265

Table 2: Summary statistics of daily percentage rates of return for emerging markets: number of observations (size), annualized mean (mean), annualized standard deviation (stdev), skewness (skew) and kurtosis (kurt).

4.2 Testing individual markets

The test statistics $M_1(k)$, $M_2(k)$, $R_1(k)$, $R_2(k)$ and $S_1(k)$ were computed using the daily returns of the 46 indices. Four lags were considered, k = 2, 5, 10, 20, corresponding to two days, one week, two weeks and one month calendar periods, respectively. Lo and MacKinlay (1988) and Wright (2000) tests are individual tests in which the random walk hypothesis is rejected if the test statistic is rejected for any of the pre-defined values of k. Chow and Denning (1993) argue that performing individual tests using several values of k may lead to an over rejection of the null-hypothesis above the nominal level of significance. To overcome the size distortion of individual tests they suggest a joint version of Lo and MacKinlay (1988) test in which the decision concerning the null hypothesis is taken on the basis of the maximum absolute value of the vector of test statistics. For instance, if individual tests are evaluated using m lags k_i, i=1,...,m, then the joint test statistics are

$$\begin{aligned} M_1' &= \max_{i=1,\ldots,m} |M_1(k_i)|, \\ M_2' &= \max_{i=1,\ldots,m} |M_2(k_i)|. \end{aligned} \tag{14}$$

Chow and Denning (1993) show that M_1' and M_2' follow a studentized maximum modulus distribution with m and T degrees of freedom. When T is large the critical value is obtained from the $[1-(\alpha^*/2)]$th percentile of the normal distribution, where $\alpha^* = 1-(1-\alpha)^{1/m}$ and α is the expected level of significance. In the spirit of Chow and Denning (1993), analogous joint tests can be devised for rank- and sign-based variance ratio tests

$$\begin{aligned} R_1' &= \max_{i=1,\ldots,m} |R_1(k_i)|, \\ R_2' &= \max_{i=1,\ldots,m} |R_2(k_i)|, \\ S_1' &= \max_{i=1,\ldots,m} |S_1(k_i)|. \end{aligned} \tag{15}$$

The sampling distribution and critical values of the joint tests R_1', R_2' and S_1' can be derived from simulation in the same fashion as they are obtained for the individual tests $R_1(k)$, $R_2(k)$ and $S_1(k)$. Through Monte Carlo simulations Belaire-Franch and Contreras (2004) show that these tests have good size and power properties against several stochastic processes. In this study, the critical values for these tests were simulated through 10,000 replications.

In order to understand how the clustering pattern evolves in the period covered by our data, we analyze three 5-years sub-samples covering the periods from 1995 to 1999, from

	1995-1999					2000-2004					2005-2009				
	M_1'	M_2'	R_1'	R_2'	S_1'	M_1'	M_2'	R_1'	R_2'	S_1'	M_1'	M_2'	R_1'	R_2'	S_1'
Australia											•		•	•	•
Austria	•		•	•	•										
Belgium	•	•	•	•	•	•					•				
Canada	•		•	•	•						•		•	•	
Denmark			•	•	•										
Finland	•		•	•	•										
France						•		•	•		•			•	
Germany															
Greece	•	•	•	•	•	•	•	•	•	•	•				•
Hong Kong				•											
Ireland	•		•	•	•										
Italy															
Japan								•		•					
Netherlands															
New Zealand			•	•	•			•	•						
Norway	•		•	•											
Portugal	•	•	•	•	•	•			•						
Singapore	•	•	•	•	•										
Spain	•	•	•	•	•										
Sweden										•			•	•	
Switzerland					•						•				
U. Kingdom	•	•	•	•		•		•	•		•		•	•	
U. States								•		•	•		•	•	•

Table 3: Results of the conventional, rank-based and sign-based variance ratio tests for developed markets, and for three different periods: 1995-1999, 2000-2004 and 2005-2009. A bullet (•) indicates that the random walk hypothesis was rejected with a statistical significance of 5%.

	1995-1999					2000-2004					2005-2009				
	M_1'	M_2'	R_1'	R_2'	S_1'	M_1'	M_2'	R_1'	R_2'	S_1'	M_1'	M_2'	R_1'	R_2'	S_1'
Argentina	•	•	•	•	•	•		•	•						
Brazil	•		•	•	•										
Chile	•	•	•	•	•	•	•	•	•	•	•		•	•	•
China	•	•	•	•	•			•	•						
Colombia	•	•	•	•	•	•	•	•	•	•	•		•	•	•
Czech Rep.	•	•	•	•	•					•					
Egypt	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Hungary	•		•	•	•						•		•	•	•
India	•	•	•	•	•	•		•	•	•	•		•	•	•
Indonesia	•	•	•	•	•	•		•	•	•	•	•	•	•	•
Israel	•		•	•	•			•							•
Korea	•	•	•	•	•										
Malaysia			•	•	•	•	•	•	•		•	•	•	•	•
Mexico	•		•	•	•	•	•	•	•	•			•	•	•
Morocco	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Peru	•	•	•	•	•								•	•	
Philippines	•	•	•	•	•	•	•	•	•	•	•	•	•	•	
Poland	•	•	•	•	•										
Russia	•	•	•	•	•					•	•				
South Africa	•		•	•	•	•	•	•	•	•					
Taiwan															
Thailand	•	•	•	•	•	•	•	•	•	•					
Turkey	•		•	•	•										

Table 4: Results of the conventional, rank-based and sign-based variance ratio tests for emerging markets, and for three different periods: 1995-1999, 2000-2004 and 2005-2009. A bullet (•) indicates that the random walk hypothesis was rejected with a statistical significance of 5%.

2000 to 2004, and from 2005 to 2009.[3] The results of testing the random walk hypothesis with statistics M_1', M_2', R_1', R_2' and S_1', for developed and emerging markets are shown in Tables 3 and 4, respectively. For a given test statistic and market, a bullet (•) indicates that the null hypothesis that the return series follows a random walk is rejected with a statistical significance of 5%. Tables 3 and 4 provide evidence for the conventional belief that index returns in emerging markets are typically more predictable that those in developed markets, as a result of the lower trading volumes and liquidity, and higher levels of regulatory restrictions. On the other hand, these results do not substantiate Griffin et al. (2010). This study found small differences between developed and emerging markets, for both stocks and portfolios, using individual Lo and MacKinlay (1988) variance ratio statistics covering the period from 1994 through 2005.

Nevertheless, some markets do not conform to this pattern. For instance, in the period 1995-1999 all test statistics reject the null hypothesis in the developed markets of Belgium, Greece, Portugal, Singapore and Spain, providing strong evidence in favor of predictability in this period. On the other hand, all tests fail to reject the null hypothesis for the market Taiwan, which is included in the emerging markets group. Also, from 2000 to 2009, all tests fail to reject the null for the market of Korea. This is no surprise since these markets have developed past the emerging market phase, despite the classification as emerging by MSCI.

The evolution of the test results across the three periods suggests a decrease of the predictability of returns in the past years. In most developed markets and many emerging markets the number of tests that fail to reject the random walk hypothesis decreases over time. The number of developed markets in which all tests fail to reject the null hypothesis increased from 8 in the first period to 14 in the most recent period. In the emerging markets group, this figure increased from 1 in the period 1995-1999 to 10 in the period 2005-2009. Remarkably, in the most recent period several tests reject the null in mature markets such as Australia, Canada, United Kingdom, and the United States.

4.3 Multidimensional scaling maps

Figure 3 shows 2-dimensional scaling maps for the three subperiods. In order to better visualize and interpret similarities among equity markets, we removed outliers which exhibited very large absolute values of the test statistics. In particular, we dropped the markets of

[3] Because the sampling theory of variance ratio tests is based on asymptotic approximations, a minimum number of observations in each subperiod is necessary. This division results in around 1250 observations in each period, a sample size that guarantees reasonably high power of the joint variance ratio tests (Belaire-Franch and Contreras, 2004).

Morocco and Colombia from the clustering analysis in subperiods 1995-1999, 2000-2004 and 2005-2009, the market of Chile in subperiods 2000-2004 and 2005-2009, and the markets of Egypt and Malaysia in subperiod 2005-2009.

The plot on the top of Figure 3 shows the multidimensional scaling (MDS) map for the first period: 1995-1999. On the right hand side of the map, standing at larger values on the horizontal axis (*Dimension* 1), one can identify a cluster containing Canada and the United States, several developed markets from western Europe (Austria, Denmark, Finland, France, Germany, Italy, Netherlands, Norway, Spain, Sweden, Switzerland and the United Kingdom) and the most developed markets from the Pacific Rim (Australia, Hong-Kong, Korea, Japan, New Zealand and Taiwan). The comparison of this cluster with the leftmost panels in Tables 3 and 4 provides an interesting result. While this cluster contains most developed markets in which *at least one* test statistic failed to reject the null hypothesis that the return series conforms to a random walk, it only contains the single emerging market in which *all* test statistics failed to reject the null (i.e., Taiwan). The exceptions to this pattern are the markets of Spain, for which all tests reject the null, and Ireland which stands between the main cluster of developed markets and the remaining developed markets.

In contrast to the developed markets, many emerging markets in which at least one test statistic fails to reject the null (i.e., Brazil, Hungary, Israel, Malaysia, Mexico, South Africa and Turkey) are displaced from this cluster. On the other hand, while all test statistics fail to reject the null in the developed market of Spain, nevertheless, it is grouped with the cluster of developed markets. Therefore, it appears that this procedure can capture similarities between markets with similar levels of development that single variance ratio tests fail to detect. On the left of this cluster, scattered more or less evenly across lower values of Dimension 1, one can find most emerging markets and four smaller developed markets (Belgium, Greece, Portugal and Singapore). Larger values of Dimension 1 appear to be directly related to lower levels of stock market predictability and, in fact, this dimension explains 86.7% of the total variance of the scaled data.

The plot in the middle of Figure 3 shows the MDS map for the period 2000-2004. The visual inspection of this map reveals a large cluster containing all developed markets with the exception of Greece. Corroborating this observation, Table 3 shows that Greece is the only developed country in which all test statistics failed to reject the null hypothesis that the return series is a random walk. We can find eight emerging markets with Dimension 1 values within the range of the developed markets (Brazil, China, Hungary, Korea, Poland, Russia, Turkey and Taiwan). Interestingly, while exactly the same tests reject the null hypothesis

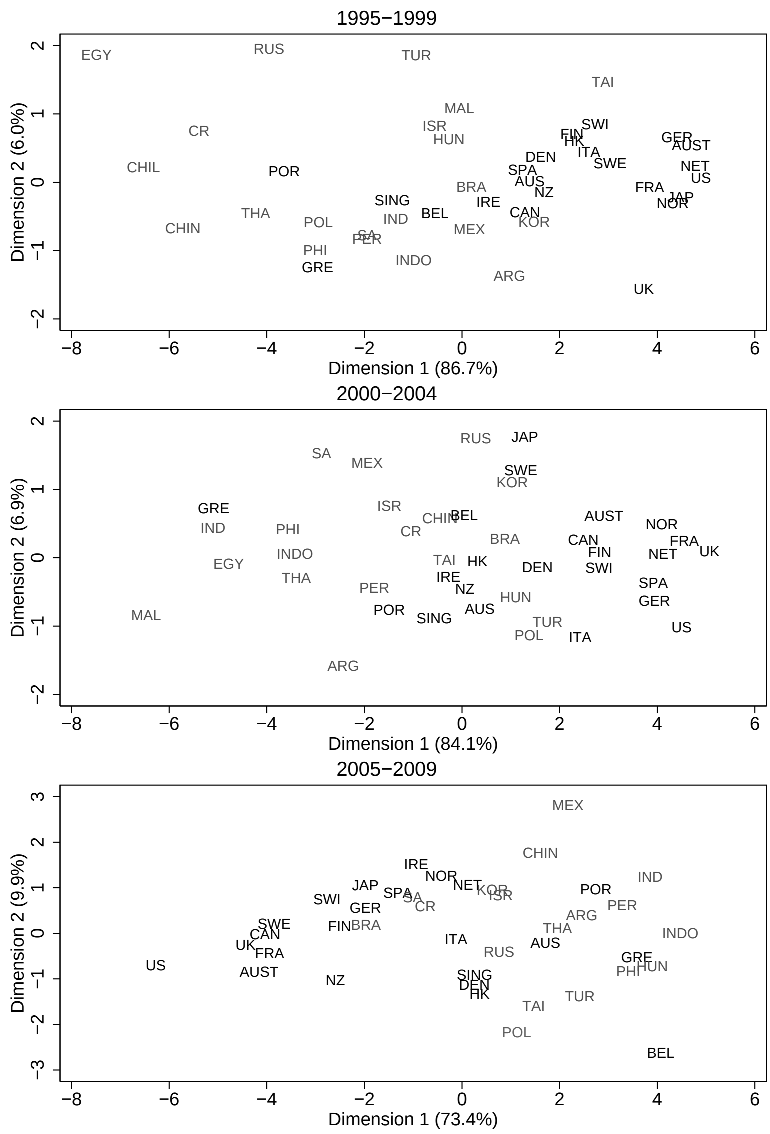

Figure 3: Two-dimensional maps of global stock markets by metric multidimensional scaling.

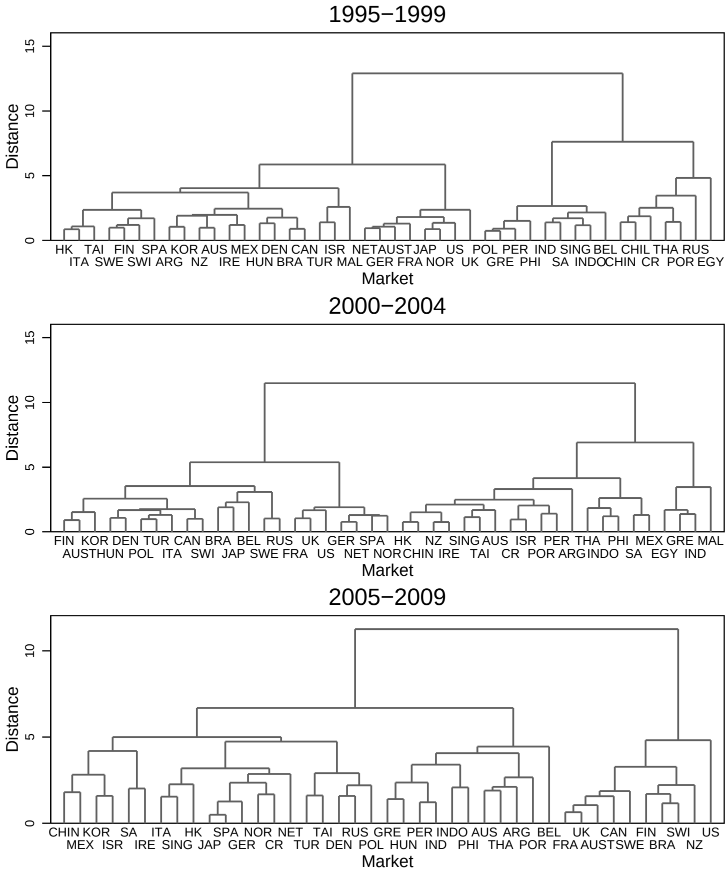

Figure 4: Dendrograms for Euclidean distances between international stock markets.

in the markets of France and Argentina, the former is found in the cluster of developed markets and the latter is found at a considerable distance from this cluster. Finally, the plot in the bottom of Figure 3 shows the MDS map for the most recent period. Apart from a reflection about the vertical axis, this map reveals a clustering pattern resembling those of the previous periods. A cluster containing most developed markets can be found at negative values of Dimension 1. As in the first period, the three European markets of Belgium, Greece and Portugal stand in the cluster of emerging markets. Interestingly, the market of the United States is the farthermost from the emerging markets cluster, despite having four tests rejecting the null during this period. Among the cluster of developed markets one can find the emerging markets of Brazil, Czech Republic and South Africa.

Period	#clusters	Duda-Hart Je(2)/Je(1)	pseudo-t^2	Clusters 1	2	3	4	5
1995-1999	1	0.4108	60.24	44				
	2	0.3945	21.48	28	16			
	3	0.5650	20.02	28	9	7		
	4	0.5133	4.74	20	9	8	7	
	5	0.7229	6.90	20	9	8	6	1
2000-2004	1	0.4465	50.83	43				
	2	0.5276	17.01	22	21			
	3	0.5685	15.18	22	17	4		
	4	0.5829	10.74	17	15	7	4	
	5	0.6343	7.49	15	12	7	5	4
2005-2009	1	0.5448	32.59	41				
	2	0.6494	15.65	31	10			
	3	0.7792	5.10	20	11	10		
	4	0.6732	3.88	14	11	10	6	
	5	0.5844	8.53	14	11	9	6	1

Table 5: Cluster solutions for Duda-Hart Je(2)/Je(1) index. The values in the rightmost cells are the number of markets for each cluster solution.

4.4 Dendrogram analysis

Figure 4 shows the dendrograms for periods 1995-1999, 2000-2004, and 2005-2009 obtained by the complete linkage method, which minimizes the maximum distance between equity markets in the same group. We have computed the Duda-Hart Je(2)/Je(1) indices and the associated pseudo-t^2 statistics to determine the optimal cluster solutions, as shown in Table 5. The appropriate number of clusters is determined by the largest Duda-Hart Je(2)/Je(1) values. The results in Table 5 suggest five clusters for periods 1995-1999 and 2000-2004, and three clusters for period 2005-2009.

Table 6 shows how the stock markets are grouped in these clusters. In period 1995-1999, cluster 1 includes 11 developed markets (Austria, Canada, Denmark, Finland, Hong-Kong, Ireland, Italy, New Zealand, Spain, Sweden and Switzerland) and nine emerging markets (Argentina, Brazil, Hungary, Israel, Korea, Malaysia, Mexico, Taiwan and Turkey). In cluster 2 we can find seven emerging markets (India, Indonesia, Peru, Philippines, Poland and South Africa) and three developed markets (Belgium, Greece and Singapore). Cluster 3 includes the indices with the largest market capitalizations, such as the United States, Japan, United Kingdom, Germany and France, together with the Netherlands and Norway. Cluster 4 contains five emerging markets (Chile, China, Czech Republic, Russia and Thailand) and

	1995 - 1999
cluster 1	Austria Canada Denmark Finland Hong-Kong Ireland Italy New Zealand Spain Sweden Switzerland Argentina Brazil Hungary Israel Korea Malaysia Mexico Taiwan Turkey
cluster 2	Belgium Greece Singapore India Indonesia Peru Philippines Poland S. Africa
cluster 3	France Germany Japan Netherlands Norway United Kingdom United States
cluster 4	Portugal Chile China Czech R. Russia Thailand
cluster 5	Egypt
	2000 - 2004
cluster 1	Australia Belgium Canada Denmark Finland Italy Japan Sweden Switzerland Brazil Hungary Korea Poland Russia Turkey
cluster 2	Austria Ireland Hong-Kong New Zealand Portugal Singapore Argentina China Czech R. Israel Peru Taiwan
cluster 3	France Germany Norway Netherlands Spain United Kingdom United States
cluster 4	Indonesia Mexico Philippines S. Africa Thailand
cluster 5	Greece Egypt India Malaysia
	2005 - 2009
cluster 1	Australia Canada Finland France New Zealand Sweden Switzerland United Kingdom United States Brazil
cluster 2	Austria Belgium Greece Portugal Argentina Hungary India Indonesia Peru Philippines Thailand
cluster 3	Denmark Germany Hong Kong Ireland Italy Japan Netherlands Norway Singapore Spain China Czech R. Israel Korea Mexico Poland S. Africa Taiwan Turkey

Table 6: Distribution of stock markets in the cluster solutions identified with the Duda-Hart stopping-rule index.

one developed market (Portugal). Finally, cluster 5 is formed by the market of Egypt. Note that, in the MDS map corresponding to this period this market can be found at a large distance from all other markets.

The clustering results over the period 2000-2004 suggest again five distinct groups. Cluster 1 includes nine developed markets (Australia, Belgium, Canada, Denmark, Finland, Italy, Japan, Sweden and Switzerland) and six emerging markets (Brazil, Hungary, Korea, Poland, Russia and Turkey). In cluster 2 we can find six emerging markets (Argentina, Czech Republic, China, Israel, Peru, Taiwan), three developed European markets (Austria, Ireland and Portugal) and three developed Pacific Rim markets (Hong-Kong, New Zealand and Singapore). Again, cluster 3 groups many of the markets with the largest market capitalizations. Interestingly, Japan is no longer included in this group. This is consistent with the corresponding MDS map, in which the market of Japan is found at a reasonable distance from other large capitalization markets. Cluster 4 only contains emerging markets (Indonesia, Mexico, Philippines, South Africa and Thailand), and cluster 5 is formed by the markets with lowest Dimension 1 values in the MDS map: Egypt, India, Malaysia and Greece.

In contrast to the previous periods, the cluster solution for period 2005-2009 indicates a separation of equity markets into three clusters. One cluster is formed by most developed markets (Australia, Canada, Finland, France, New Zealand, Sweden, Switzerland, United Kingdom and United States) and Brazil. A second cluster is formed by seven emerging markets (Argentina, Hungary, India, Indonesia, Peru, Philippines and Thailand) and four developed European markets (Austria, Belgium, Greece and Portugal). A third cluster is formed by the remaining developed and emerging markets.

5 Conclusions

In this paper, we introduced a new distance measure for clustering time series which is based on variance ratio statistics. The proposed metric attempts to gauge the level of interdependence of time series from the point of view of return predictability. Simulation results show that a variance ratio-based metric aggregates better time series according to their serial dependence structure than a metric based on the sample autocorrelations. An empirical application of this approach to a large set of 46 international stock market returns was presented. The examination of individual variance ratio statistics provided further evidence for the conventional belief that the returns of emerging stocks markets are usually more predictable than those of their developed counterparts. It was also found that in the period covered by the data there was a trend towards greater unpredictability of returns in both developed and emerging markets. Groups of markets were identified by visual inspection of two-dimensional scaling maps and by cluster solutions given by the Duda-Hart stopping-rule index. The propose metric clustered reasonably well the stock markets according to their size and level of development. It also provided interesting insights that the analysis of single variance ratios failed to capture.

While this analysis was conducted using specific variance ratio tests and pre-determined lag orders, future studies can explore alternative variance ratio tests, such as those with automatic selection of the optimal holding period (see, Kim, 2009; Charles et al., 2011). Furthermore, future studies could attempt to relate our clustering metric to factors that may affect market efficiency, such as regulatory restrictions, freedom of capital movements, and trade openness (see, Lim and Kim, 2011).

Acknowledgments

This study was supported by a grant from FCT - Fundação para a Ciência e Tecnologia.

References

Belaire-Franch, J., Contreras, D., 2004. Ranks and signs-based multiple variance ratio tests. Working paper. University of Valencia.

Berndt, D.J., Clifford, J., 1996. Finding patterns in time series: a dynamic programming approach. In: Advances in Knowledge Discovery and Data Mining (Ed.: U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy). AAAI Press/The MIT Press, Menlo Park, 229–248.

Bohte, Z.D., Cepar, D., Kosmelu, K., 1980. Clustering of time series, COMPSTAT 80, Physica-Verlag, 587-593.

Bonanno, G., Lillo, F., Mantegna, R., 2001. High-frequency cross-correlation in a set of stocks. Quantitative Finance 1, 96-104.

Caiado, J., Crato, N., Peña, D. (2006). A periodogram-based metric for time series classification. Computational Statistics & Data Analysis 50, 2668-2684.

Caiado, J., Crato, N., Peña, D. (2009). Comparison of time series with unequal length in the frequency domain. Communications in Statistics - Simulation and Computation 38, 527-540.

Caiado, J., Crato, N., 2010. Identifying common dynamic features in stock returns. Quantitative Finance 10, 797-807.

Charles, A., Darné, O., 2009. Variance-ratio tests of random walk: an overview. Journal of Economic Surveys 23, 503527.

Charles, A., Darné, O., Kim, J.H., 2011. Small sample properties of alternative tests for martingale difference hypothesis. Economics Letters 110, 151-154.

Chow, K.V., Denning, K.C., 1993. A simple multiple variance ratio test. Journal of Econometrics 58, 385-401.

Duda, R.O., Hart, P.E., 1973. Pattern Classification and Scene Analysis. Wiley, New York.

Everitt, B., Landau, S., Leese, M., 2001. Cluster Analysis, 4th Ed. Edward Arnold, London.

Galeano, P., Peña, D. (2000). Multivariate analysis in vector time series. Resenhas 4, 383-404.

De Gregorio, A., Iacus, S.M., 2008. Clustering of discretely observed diffusion processes. Computational Statistics & Data Analysis 54, 598-606.

Griffin, J.M., Kelly, P.J., Nardari, F., 2010. Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. Review of Financial Studies 23, 32253277.

Johnson, R., Wichern, D., 2007. Applied Multivariate Statistical Analysis, 6th Ed. Prentice-Hall, New Jersey.

Kalpakis, K., Gada, D., Puttagunta, V., 2001. Distance measures for effective clustering of ARIMA time-series. Proceedings of the IEEE International Conference on Data Mining, San Jose, 273-280.

Kim, J.H., 2009. Automatic variance ratio test under conditional heteroskedasticty. Finance Research Letters 6, 179-185.

Kim, E.H., Singal, V., 2000. Stock market openings: experience of emerging economies. The Journal of Business 73(1), 25-66.

J.H. Kim and A. Shamsuddin, 2008. Are Asian stock markets efficient Evidence from new multiple variance ratio tests. Journal of Empirical Finance 15, 518-532.

Hoque, H.A.A.B., Kim, J.H., Pyun, C.S., 2007. A comparison of variance ratio tests of random walk: A case of Asian emerging stock markets. International Review of Economics and Finance 16, 488–502.

Lim, K.-P., Brooks, R., 2011. The evolution of stock market efficiency over time: a survey of the empirical literature. Journal of Economic Surveys 25, 69108.

Lim, K.-P., Kim, J.H., 2011. Trade openness and the informational efficiency of emerging stock markets. Economic Modelling 28, 2228-2238.

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

Lo, A.W., MacKinlay, A.C., 1989. The size and power of the variance ratio test in finite samples: A Monte Carlo investigation. Journal of Econometrics 40(2), 203–238.

Maharaj, E.A., 1996. A significance test for classifying ARMA models. Journal of Statistical Computing and Simulation 54, 305-331.

Maharaj, E.A., 1999. Comparison and classifying of stationary multivariate time series. Pattern Recognition 32, 1129-1138.

Maharaj, E.A., 2000. Clusters of time series. Journal of Classification 17, 297-314.

Maharaj, E.A., 2002. Comparison of non-stationary time series in the frequency domain. Computational Statistics & Data Analysis 40, 131-141.

Maharaj, E.A., D'Urso, P.A., 2010. A coherence-based approach for the pattern recognition of time series. Physica A: Statistical mechanics and its Applications 389, 3516-3537.

Mantegna, R.N., 1999. Hierarchical structure in financial markets. The European Physical Journal B, 11, 193-197.

Möller-Levet, C. S., Klawonn, F., Cho, K. and Wolkenhauer, O. (2003). Fuzzy clustering of short time-series and unevenly distributed sampling points In: Advances in Intelligent Data Analysis V, volume 2810 of Lecture Notes in Computer Science, Springer Verlag, 330-340.

Piccolo, D., 1990. A distance measure for classifying ARIMA models. Journal of Time Series Analysis 11, 152-164.

Savvides, A., Promponas, V. and Fokianos, K., 2008. Clustering of biological time series by cepstral coefficients based distances. Pattern Recognition 41, 2398-2412.

Smith, G., 2009. Martingales in European emerging stock markets: Size, liquidity and market quality. The European Journal of Finance 15, 249-262.

Tong, H., Dabas, P., 1990. Cluster of time series models: an example. Journal of Applied Statistics 17, 187-198.

Wang, K., Gasser, T., 1997. Alignment of curves by dynamic time warping. Annals of Statistics 25, 1251-1276.

Wright, J.H., 2000. Alternative variance-ratio tests using ranks and signs. Journal of Business and Economics Statistics 18(1), 1-9.