Identifying common spectral and asymmetric features in stock returns

Caiado, Jorge and Crato, Nuno

December 2007

Identifying common spectral and asymmetric features in stock returns

Jorge Caiado[a] Nuno Crato[b]

[a]Department of Economics and Management, School of Business Administration, Polytechnic Institute of Setúbal, and CEMAPRE, Campus do IPS, Estefanilha, 2914-503 Setúbal, Portugal. Tel.: +351 265 709 438. Fax: +351 265 709 301. E-mail: jcaiado@esce.ips.pt

[b]Department of Mathematics, School of Economics and Business, Technical University of Lisbon, and CEMAPRE, Rua do Quelhas 6, 1200-781 Lisboa, Portugal. Tel.: +351 213 925 846. E-mail: ncrato@iseg.utl.pt

Abstract: This paper proposes spectral and asymmetric-volatility based methods for cluster analysis of stock returns. Using the information about both the periodogram of the squared returns and the estimated parameters in the TARCH equation, we compute a distance matrix for the stock returns. Clusters are formed by looking to the hierarchical structure tree (or dendrogram) and the computed principal coordinates. We employ these techniques to investigate the similarities and dissimilarities between the "blue-chip" stocks used to compute the Dow Jones Industrial Average (DJIA) index. For reference, we investigate also the similarities among stock returns by mean and squared correlation methods.

Keywords: Asymmetric effects; Cluster analysis; DJIA stock returns; Periodogram; Threshold ARCH model; Volatility.

1. Introduction

Cluster analysis of financial time series plays an important role in several areas of application. In stock markets, the examination of mean and variance correlations between asset returns can be useful for portfolio diversification and risk management purposes. In international equity markets, we may be interested in identifying similarities in index returns and volatilities for grouping countries. The existence of asymmetric cross-correlations and dependences in asset returns is also of interest for many financial researchers.

Many time-varying volatility models have been proposed to capture the asymmetric volatility effects in asset returns. These include the common univariate asymmetric models of Nelson (1991), Engle and Ng (1993), Glosten, Jagannathan and Runkle (1993) and Zakoian (1994), the multivariate generalized autoregressive conditionally heteroskedasticity (GARCH) models of Engle and Kroner (1995) and Kroner and Ng (1998), and the asymmetric dynamic autoregressive conditional correlation model of Capiello, Engle and Sheppard (2006).

Two different types of asymmetries in asset returns have been investigated in the econometric and finance literature. The first type is concerned with the leverage effect sometimes observed in individual stock returns (see Christie, 1982 and Schwert, 1989). Analyzing the conditional variance, most studies find a higher persistence for negative returns than for positive returns (see, for instance, Engle and Ng, 1993, and Glosten, Jagannathan and Runkle, 1993). The second type of asymmetry is found on the comovements of stock returns. For example, Yu and Wu (2001) analyzed the asymmetric cross-correlation in stock returns relative to economic factors. Ang and Chen (2002) developed statistics for comparing and testing asymmetries in conditional correlations between U.S. stocks. Most empirical evidence shows an asymmetric pattern in the dependence of stock returns in the sense that stock returns exhibit higher correlations in periods of market downturns than in periods of upturns. For a review of the research literature on asymmetric volatility, see the surveys by Bollerselev, Chou and Kroner (1992), Kroner and Ng (1998) and Bekaert and Wu (2000).

Many existing statistical methods for analysis of multiple asset returns use multivariate volatility models imposing conditions on the covariance matrix that are hard to apply. To avoid these problems, three types of multivariate statistical techniques have been used for analysing the structure of asset returns comovements. One is the principal component analysis (PCA) that is concerned with the covariance structure of asset returns and can be used in dimension reduction. The second is the factor model for asset returns that uses multiple time series to describe the common factors of returns (see Zivot and Wang, 2003 and Tsay, 2005 for further discussion). The third is the identification of similarities in asset return volatilities using cluster analysis (see, for instance, Bonanno, Caldarelli, Lillo, Miccieché, Vandewalle and Mantegna, 2004).

A fundamental problem in clustering of financial time series is the choice of a relevant metric. Mantegna (1999), Bonanno, Lillo and Mantegna (2001), among others, used the Pearson correlation coefficient as similarity measure of a pair of stock returns. Although this metric can be useful to ascertain the structure of stock returns movements, it does not take into account the stochastic volatility dependence of the processes and cannot be used for comparison and grouping stocks with unequal sample sizes. The later is a common problem of most existing nonparametric-based metrics for cluster analysis of economic and financial time series.

In this paper, we introduce a distance measure between the threshold autoregressive conditionally heteroskedastic (TARCH) parameters of the return series. In order to also capture the spectral behavior of the time series, we suggest combining the proposed statistic with a periodogram distance measure for the squared returns. In order to summarize and better interpret the results, we suggest using a hierarchical clustering tree and a multidimensional scaling map to explore the existence of clusters.

We apply these steps to investigate the similarities and dissimilarities among the "blue-chip" stocks of the Dow Jones Industrial Average (DJIA) index. For reference with a well-known method, we also investigate the similarities and dissimilarities among level and squared returns by using correlation measures.

The remaining sections are organized as follows. Section 2 provides the asymmetric-volatility and spectral based methods for clustering asset returns. Section 3 describes the data. Section 4 presents the empirical findings on the analyzed data. Section 5 compares these findings with those obtained by squared returns correlations. Section 6 summarizes and concludes.

2. Asymmetric-volatility and spectral based distances

Glosten, Jagannathan and Runkle (1993) and Zakoian (1994) introduced independently the Threshold ARCH model to allow for asymmetric shocks to volatility. The simple TARCH(1,1) model assumes the form

$$\varepsilon_t = z_t\sigma_t, \tag{1}$$
$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 d_{t-1}, \tag{2}$$

where $\{z_t\}$ is a sequence of independent and identically distributed random variables with zero mean and unit variance, $d_t = 1$ if ε_t is negative, and $d_t = 0$ otherwise. In this model, volatility tends to rise with the "bad news" ($\varepsilon_{t-1} < 0$) and to fall with the "good news" ($\varepsilon_{t-1} < 0$). Good news has an impact of α while bad news has an impact of $\alpha + \gamma$. If $\gamma > 0$ then the leverage effect exists. If $\gamma \neq 0$, the shock is asymmetric, and if $\gamma = 0$, the shock is symmetric. The persistence of shocks to volatility is given by $\alpha + \beta + \gamma/2$. Nelson (1991) proposed also an heteroskedasticity model to incorporate the asymmetric effects between positive and negative stock returns, called the exponential GARCH (or EGARCH) model, in which the leverage effect is exponential rather than quadratic. To capture all the skewness and excess kurtosis in the volatility processes with asymmetric distributions, Nelson (1991) suggested a "fat-tailed" distribution, the generalized error distribution (GED), with density function given by

$$f(z) = \frac{v\exp\left[-0.5\left|z/\lambda\right|^v\right]}{\lambda 2^{(1+1/v)}\Gamma(1/v)}, 0 < v \leq \infty, -\infty < z < +\infty \tag{3}$$

where v is the tail-tickness parameter, $\Gamma(\cdot)$ is the gamma function, and

$$\lambda = \left[\frac{2^{(-2/v)}\Gamma(1/v)}{\Gamma(3/v)}\right]^{0.5}. \tag{4}$$

When $v = 2$, $\{z_t\}$ is normally distributed, and is fat-tailed distributed if $v < 2$. For $v > 2$, it has thinner tails distribution (for example, for $v = +\infty$, it has a uniform distribution on the interval $[-\sqrt{3}, \sqrt{3}]$).

We now introduce a distance measure for clustering time series with similar asymmetric volatility effects. Let $r_{x,t} = \log P_{x,t} - \log P_{x,t-1}$ denote the continuously compounded return of an asset x from time $t-1$ to t ($r_{y,t}$ is similarly defined for asset y). Suppose we fit a common TARCH(1,1) model to both time series by the method of maximum likelihoods assuming GED innovations. Let $T_x^G = (\widehat{\alpha}_x, \widehat{\beta}_x, \widehat{\gamma}_x, \widehat{v}_x)'$ and $T_y^G = (\widehat{\alpha}_y, \widehat{\beta}_y, \widehat{\gamma}_y, \widehat{g}_y)'$ be the vectors of the estimated ARCH, GARCH, leverage effect and tail-tickness parameters, respectively, with the estimated covariance matrices given by V_x^G and V_y^G, respectively. A Mahalanobis-like distance between the asymmetric features of the volatilities (TARCH-based distance) of the return series $r_{x,t}$ and $r_{y,t}$ can be defined by

$$d_{TARCH}(x, y) = \sqrt{(T_x^G - T_y^G)'\Omega^{-1}(T_x^G - T_y^G)}, \tag{5}$$

where $\Omega = V_x^G + V_y^G$. This measure takes into account the information about the asymmetric structure of the time series volatilities and solves the problem of of unequal lengths.

We can used alternative methods based on the periodogram ordinates and the autocorrelations lags of the squared returns. The autocorrelation function and the spectrum for the squared return series provide useful information about the time series behavior in terms of the ARCH effects. Let $P_x^{SR}(\omega_j) = n^{-1}|\sum_{t=1}^{n} r_{t,x}e^{-it\omega_j}|^2$ be the periodogram of the squared return series, $r_{x,t}^2$, at frequencies $\omega_j = 2\pi j/n$, $j = 1, ..., [n/2]$ (with $[n/2]$ the largest integer less or equal to $n/2$) in the range 0 to π, and s_x^2 be the sample variance of $r_{x,t}$ (similar expression applies to asset y), the Euclidean distance between the log normalized periodograms (Caiado, Crato and Peña, 2006) of the squared returns of x and y is given by

$$d_{LNP}(x,y) = \sqrt{\sum_{j=1}^{[n/2]} \left[\log(P_x^{SR}(\omega_j)/s_x^2) - \log(P_y^{SR}(\omega_j)/s_y^2)\right]^2}, \tag{6}$$

or, using matrix notation,

$$d_{LNP}(x,y) = \sqrt{(L_x - L_y)'(L_x - L_y)}. \tag{7}$$

where L_x and L_y are the vectors of the log normalized periodogram ordinates of the squared return series, $r_{x,t}^2$ and $r_{y,t}^2$, respectively. Since the parametric features of the TARCH model are not necessary associated with all the periodogram ordinates and all the autocorrelation lags, the parametric and nonparametric approaches can be combined to take into account both the asymmetric stochastic dependence and the cyclical behavior of the return series, that is

$$d_{TARCH-LNP}(x,y) = \lambda_1\sqrt{(T_x^G - T_y^G)'\Omega^{-1}(T_x^G - T_y^G)} + \lambda_2\sqrt{(L_x - L_y)'(L_x - L_y)}. \tag{8}$$

where $\lambda_i, i = 1, 2$ are normalizing constants. The distance measures (5) and (8) fulfil the usual properties of a metric (except the triangle inequality): (i) $d(x,y)$ is asymptotically zero for independent time series generated by the same DGP; (ii) $d(x,y) \geq 0$; and (iii) $d(x,y) = d(y,x)$.

3. Data description

We consider data of the 30 "blue-chip" US daily stocks used to compute the Dow Jones Industrial Average (DJIA) index for the period from June 1990, 11 to September 2006, 12 (4100 daily observations), as shown in Table 1. This data was obtained from Yahoo Finance (http://finance.yahoo.com) and correspond to closing prices adjusted for dividends and splits.

Table 2 presents the summary statistics (mean, standard deviations, skewness, kurtosis, and Ljung-Box test statistic for serial correlation) for daily stock returns. Hewlett-Packard, Inter-tel, Microsoft, and AT&T (technology companies), Boeing, Caterpillar, and Honeywell (industrial goods), Walt Disney, Home Depot, and Mcdonalds (services),

Table 1
Stocks used to compute the Dow Jones Industrial Average (DJIA) Index

Stock	Code	Sector	Stock	Code	Sector
Alcoa Inc.	AA	Basic materials	Johnson & Johnson	JNJ	Healthcare
American Int. Group	AIG	Financial	JP Morgan Chase	JPM	Financial
American Express	AXP	Financial	Coca-Cola	KO	Consumer goods
Boeing Co.	BA	Industrial goods	McDonalds's	MCD	Services
Caterpillar Inc.	CAT	Financial	3M Co.	MMM	Conglomerates
Citigroup Inc.	CIT	Industrial goods	Altria Group	MO	Consumer goods
El Dupont	DD	Basic materials	Merck & Co.	MRK	Healthcare
Walt Disney	DIS	Services	Microsoft Corp.	MSFT	Technology
General Electric	GE	Industrial goods	Pfizer Inc.	PFE	Healthcare
General Motors	GM	Consumer goods	Procter & Gamble	PG	Consumer goods
Home Depot	HD	Services	AT&T Inc.	T	Technology
Honeywell	HON	Industrial goods	United Technologies	UTX	Conglomerates
Hewlett-Packard	HPQ	Technology	Verizon Communic.	VZ	Technology
Int. Business Machin.	IBM	Technology	Walt-Mart Stores	WMT	Services
Inter-tel Inc.	INTC	Technology	Exxon Mobile CP	XOM	Basic materials

Johnson & Johnson, Merck, and Pfizer (healthcare), Coca-cola, Altria, and Procter & Gamble (consumer goods) exhibit a negative skewness, which show the distribution of those returns have long left tails. Moreover, the higher negative skewness coefficients correspond to returns series (BA, HD, INTC, MO, MRK, PG, UTX) with higher excess of kurtosis. All financial companies and basic materials companies have a positive skewness coefficient. There are no significant autocorrelations up to order 20 in the returns for companies BA, CAT, DD, DIS, GE, GM, HON, HPQ, IBM, JPM, and MCD.

In Table 3 we present the estimation results of TARCH(1,1) models for DJIA stock returns with GED innovations, including diagnostic tests for residual and squared residuals. The estimated coefficients are statistically significant for all stocks except the ARCH estimates for CAT, DIS, GE and MRK, and the leverage-effect for INTC and MMM, which are not significant at conventional levels. The distribution of the innovation series is fat-tailed for all stocks. As expected, the persistent estimates for all the asymmetric models are very close to one. This extreme persistence in the conditional variance is very common in many empirical application using high frequency data (see Bollerselev, Chou and Kroner, 1992, and Kroner and Ng, 1998).

The Lagrange multiplier test statistic show evidence of no serial correlation in the squared residuals up to order 20 for all stocks except CAT, MCD and VZ. In terms of the mean equation, the Ljung-Box test statistic do not reject the null hypothesis of no serial correlation in the residuals for all stocks except AIG, JNJ, PFE, UTX, VZ, and XOM.

4. Cluster analysis using the TARCH-LNP based distance

Cluster analysis of time series attempts to determine groups (or clusters) of objects in a multivariate data set. The most commonly used partition clustering method is based in

Table 2
Summary statistics for Dow Jones Industrial Average (DJIA) stock returns

Stock	Mean×100	Std. dev.×100	Skewness	Kurtosis	$Q(20)$
AA	0.037	2.043	0.226	5.750	32.4**
AIG	0.051	1.696	0.131	6.247	56.8*
AXP	0.066	2.108	0.291	8.966	35.1**
BA	0.030	1.951	-0.535	10.666	26.2
CAT	0.059	1.997	-0.032	6.019	23.8
CIT	0.080	2.135	0.021	7.509	33.2**
DD	0.030	1.735	0.073	5.890	28.2
DIS	0.029	1.987	-0.081	10.241	23.3
GE	0.053	1.646	0.042	7.062	29.8
GM	0.011	2.113	0.095	6.499	27.6
HD	0.064	2.187	-0.952	19.773	53.7*
HON	0.044	2.104	-0.152	15.327	14.5
HPQ	0.055	2.614	-0.098	8.386	8.4
IBM	0.031	1.961	0.001	9.753	25.6
INTC	0.082	5.495	-0.258	11.978	370.8*
JNJ	0.058	1.534	-0.256	8.665	98.0*
JPM	0.042	2.260	0.119	8.020	27.4
KO	0.040	1.570	-0.082	7.038	42.3*
MCD	0.041	1.723	-0.063	7.055	16.3
MMM	0.042	1.475	0.028	7.143	33.9**
MO	0.060	1.927	-0.802	18.509	39.5**
MRK	0.040	1.810	-1.355	27.212	48.1*
MSFT	0.082	2.216	-0.041	7.471	21.8*
PFE	0.065	1.868	-0.135	5.349	46.7*
PG	0.052	1.612	-2.823	66.497	50.1*
T	0.034	1.762	-0.071	6.671	32.3**
UTX	0.061	1.773	-1.235	26.060	36.8**
VZ	0.024	1.707	0.126	6.976	59.3*
WMT	0.048	1.889	0.075	5.384	58.3*
XOM	0.054	1.404	0.022	5.545	69.7*

* (**) Significant at the 1% (5%) level; $Q(20)$ is the Ljung-Box statistic with 20 lags.

Table 3
Estimated TARCH(1,1) models with conditional GED innovations for DJIA stock returns

Stock	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{v}$	Persistence	$Q(20)$	$Q^2(20)$	$LM(20)$
AA	0.02403*	0.95053*	0.03220*	1.482*	0.9907	26.4	19.3	18.9
AIG	0.04141*	0.91677*	0.05873*	1.417*	0.9874	35.0**	15.6	16.3
AXP	0.01958*	0.94808*	0.06949*	1.343*	1.0024	24.2	3.2	3.2
BA	0.03346*	0.93562*	0.03709*	1.317*	0.9876	15.5	21.8	21.0
CAT	0.00340	0.98055*	0.02344*	1.320*	0.9957	21.9	36.2**	16.3
CIT	0.02722*	0.95570*	0.03781*	1.405*	1.0018	21.1	17.0	16.9
DD	0.01787*	0.96790*	0.02372*	1.466*	0.9976	15.1	16.2	16.4
DIS	0.00494	0.97643*	0.03166*	1.344*	0.9972	17.5	10.7	10.4
GE	0.00816	0.96498*	0.05153*	1.598*	0.9989	17.6	21.1	21.2
GM	0.02065*	0.94330*	0.04757*	1.380*	0.9877	23.0	13.5	13.2
HD	0.01317*	0.95588*	0.05286*	1.397*	0.9955	29.8	7.7	7.9
HON	0.04347*	0.87160*	0.11698*	1.247*	0.9736	17.7	16.5	16.3
HPQ	0.01362*	0.97216*	0.01908*	1.224*	0.9953	19.6	9.0	8.9
IBM	0.02417*	0.95046*	0.04493*	1.259*	0.9971	14.2	12.1	11.8
INTC	0.02642*	0.96920*	0.00817	0.969*	0.9997	25.7	11.2	11.0
JNJ	0.03090*	0.93535*	0.06490*	1.450*	0.9999	35.5**	26.1	26.5
JPM	0.02044*	0.95543*	0.06946*	1.418*	1.0006	27.2	15.0	14.9
KO	0.02089*	0.95719*	0.04040*	1.416*	0.9983	22.8	22.6	22.7
MCD	0.01897*	0.95870*	0.02784*	1.405*	0.9916	13.9	44.6*	45.5*
MMM	0.01216*	0.98754*	-0.00219	1.186*	0.9986	21.9	17.1	16.6
MO	0.06040*	0.88601*	0.05836*	1.098*	0.9756	16.3	3.7	4.0
MRK	0.01701	0.90773*	0.06365*	1.186*	0.9566	28.8	0.9	0.9
MSFT	0.05052*	0.92676*	0.04293*	1.316*	0.9988	10.8	6.2	6.4
PFE	0.04057*	0.93469*	0.02592**	1.468*	0.9882	31.9**	11.6	11.2
PG	0.03159*	0.94220*	0.04236*	1.336*	0.9950	26.9	2.6	2.8
T	0.03919*	0.93948*	0.03402*	1.450*	0.9957	22.1	22.4	22.7
UTX	0.02540*	0.90959*	0.10784*	1.324*	0.9889	32.2**	4.4	4.4
VZ	0.02877*	0.94453*	0.04853*	1.520*	0.9976	33.6**	41.2*	37.8*
WMT	0.02549*	0.95718*	0.03206*	1.543*	0.9987	30.2	18.9	18.2
XOM	0.03407*	0.93796*	0.03420*	1.610*	0.9891	45.8*	26.1	26.4

* (**) Significant at the 1% (5%) level; $Q(20)$ is the Ljung-Box statistic for serial correlation in the residuals up to order 20; $Q^2(20)$ is the Ljung-Box statistic for serial correlation in the squared residuals up to order 20 (McLeod and Li, 1983); $LM(20)$ is the Lagrange multiplier test statistic for ARCH effects (Engle, 1982) in the residuals up to order 20.

hierarchical classifications of the objects. In hierarchical cluster analysis, we begin with each time series being considered as a separate cluster (k clusters). In the second stage, the closest two groups are linked to form $k-1$ clusters. This process continues until the last stage in which all the time series are in the same cluster (see Everitt, Landau and Leese, 2001 for further discussion).

Figure 1 shows the cluster analysis of DJIA stock returns using a hierarchical clustering tree (or dendrogram) by complete linkage (see, e.g., Johnson and Wichern, 2002). For this purpose we used the combined TARCH-LNP distance measure (8), with weights proportional to the distance.

Figure 2 shows the multidimensional scaling map of distances constructed with the same distance measure. The multidimensional scaling is a multivariate statistical method closely related to principal coordinates analysis, and uses the information about the similarities (or dissimilarities) between the time series to construct a configuration of k points in the r-dimensional space (in this case, two dimensions). For details, see Morrison (2005). The plot can also help to identify the clusters.

The dendrogram associated with the stochastic features of returns series suggests homogenous clusters of stocks with respect to the economic sectors. We found a group formed by consumer goods companies (Coca-Cola, General Motors and Procter & Gamble), by industrial goods companies (Boeing and Citigroup), by financial companies (American Int. Group, American Express and JP Morgan Chase), by technology companies (IBM, Microsoft and AT&T), by healthcare companies (Johnson & Johnson and Pfizer) and by the Home Depot company; a group formed by basic materials companies (Alcoa, El Dupont and Exxon) and by services companies (McDonalds and Walt-Mart Stores); a group formed by a miscellaneous sector companies (Caterpiller, Walt Disney, Hewlett-Packard and 3M); a group formed by Honeywell and United Technologies; and a group formed by Altria and Merck. The Inter-Tel company was not grouped. Looking at the map of distances across the stocks, we appear to have most financial and technology companies close together, most services and basic materials companies tend to cluster together, and most consumer goods companies are close to each other and close to the financial-technology cluster at the first coordinate. The two conglomerates companies (MMM and UTX) are quite distinct at the second coordinate, and again the INTC company is a clear outlier.

5. A comparative study using cross-correlation methods

In this section, we investigate the similarities among DJIA stocks by contemporaneous cross-correlation methods.

In order to identify initial stock-returns linkages, we have computed cross-correlations both for the DJIA returns and for the squared returns (detailed results are available from the authors upon request). We found that pairs of stocks in general show lower cross-correlations in squared returns than in returns. Inter-Tel, McDonalds, Altria Group, Merck and Procter & Gamble were found to be the most independent companies in volatilities, while Inter-Tel, McDonalds and American Express were the lowest return correlated companies. Pairs formed among United Technologies-Boeing-Walt Disney, United Technologies-Boeing-Honeywell, United Technologies-General Motors, AT&T-Verizon, and

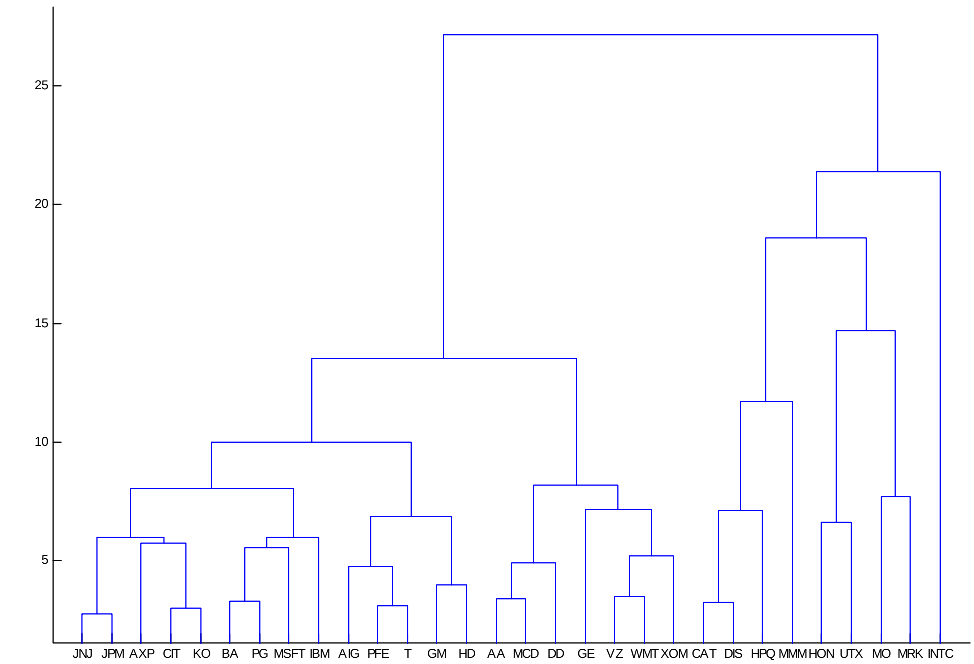

Figure 1. Dendrogram of DJIA stock returns using the TARCH-LNP distance

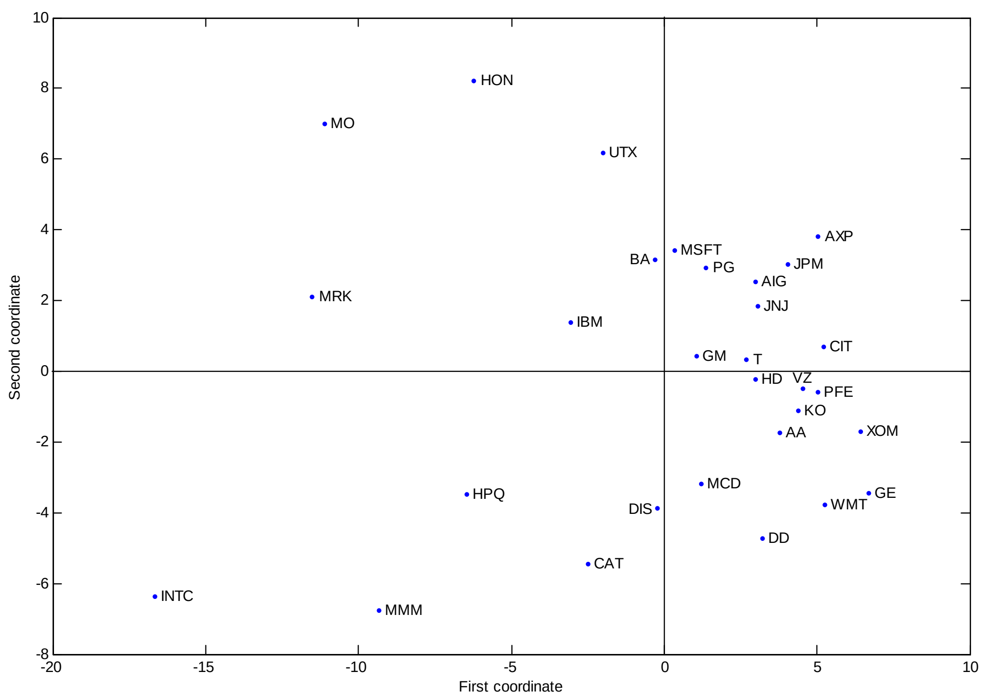

Figure 2. Multidimensional scaling of DJIA stock returns using the TARCH-LNP distance

AIG-General Electric showed the highest cross-correlations in volatilities, and pairs formed among AT&T-Verizon (communications), IBM-Hewlett-Packard-Microsoft (technology), Johnson & Johnson-Merck-Pfizer (healthcare), Walt-Mart Stores-Home Depot (services), Honeywell-United Technologies-Boeing, Citigroup-Home Depot, El Dupont-3M, and General Electric-AIG displayed the highest cross-correlations in levels.

We then used these cross correlations as a basis for a new cluster analysis. We choose a well-established correlation metric: $d(x,y) = \sqrt{2(1 - Cor(r_x, r_y)}$ (see Mantegna, 1999 and Bonnano, Lillo and Mantegna, 2001). Again, we used both the level returns and the squared returns. Figures 3, 4, 5 and 6 show the complete linkage dendrograms and the two-dimensional scaling maps for DJIA mean and squared correlation distances.

The hierarchical structure of the stocks is less clear in these dendrograms (Figures 3 and 5) than in the TARCH-LNP dendrogram (Figure 1). Therefore, trying to determine the appropriate partition is not straightforward. Only the metric scaling maps (Figures 4 and 6) suggest a linkage structure for the stocks we now describe.

For level returns, we found three groups of companies. One group is composed of healthcare companies (JNJ, MRK and PFE), consumer goods companies (KO, MO and PG), Communications companies (T and VZ) and XOM. The second group is composed of industrial goods companies (BA, HON and CIT), conglomerates companies (UTX and MMM), services companies (HD, WMT and DIS), technology companies (HPQ, IBM and MSFT) and companies CAT, DD and GM. The third is composed of financial companies (AXP, AIG, JPM) and miscellaneous sector companies (GE, AA, AXP, MCD and INTC).

For squared returns, we can see that the technology companies are close to each other in a distinct cluster. Most financial companies (AIG, AXP and JPM) tend to cluster together. The healthcare companies (JNJ, MRK and PFE) tend to group with some consumer goods (KO and PG) and basic materials (XOM and DD) companies. There is a miscellaneous sector group formed by industrial goods companies (BA and HON) and by the companies GM, UTX and DIS.

From the dendrograms and the scaling maps we note that all considered methods are able of getting meaningful company clusters. The TARCH-LNP method tends to collect most financial, technology, consumer goods and industrial goods companies in a distinct cluster and most basic materials and services companies in another distinct cluster. The mean correlation method tends to group the healthcare and consumer goods companies in a cluster, the industrial goods, conglomerates, technology and services companies in another one, and the financial companies in a third one. The squared correlation method tends to group most healthcare, basic materials and consumer goods companies in a cluster, most financial companies in another one, and the technology companies in a third one.

Clearly, the three methods provide somehow different clustering results. From the dendrograms, it is apparent that the proposed TARCH-LNP method can distinguish better among different financial time series. Its vertical links are significantly longer, which means that the horizontal cutoff can be made in a more reliable fashion, providing clearly separated clusters.

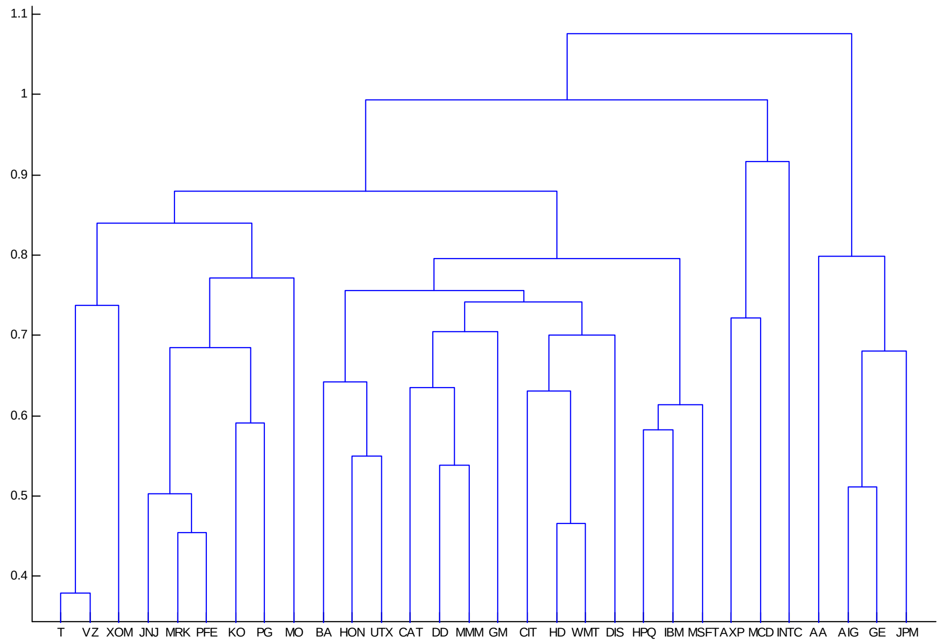

Figure 3. Dendrogram of DJIA returns using the correlation-based metric

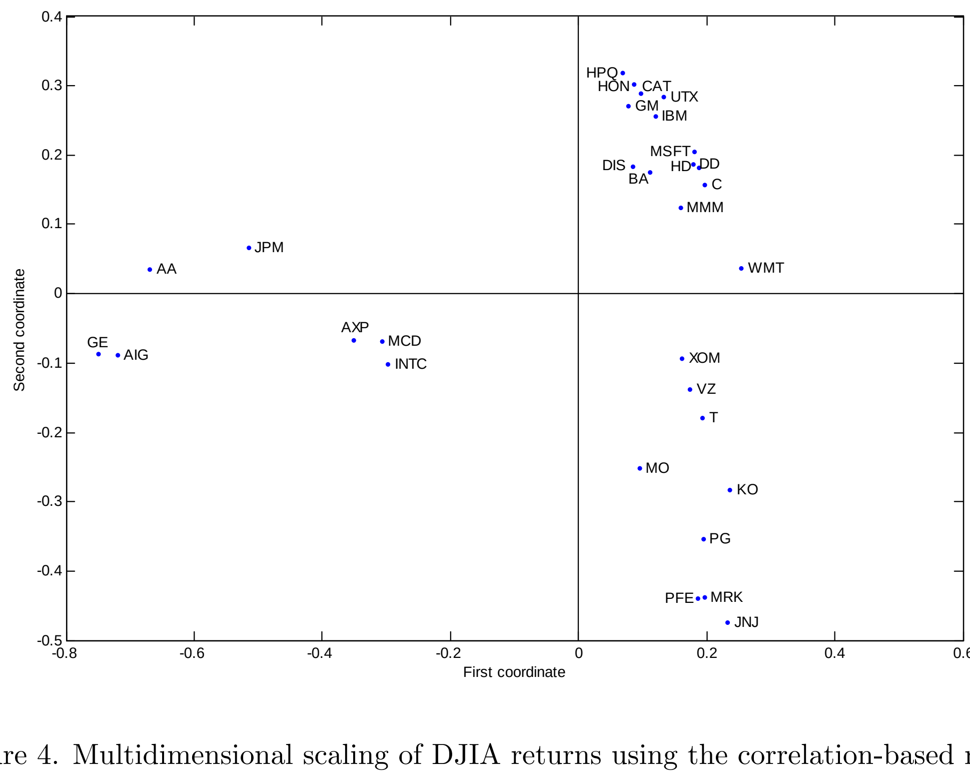

Figure 4. Multidimensional scaling of DJIA returns using the correlation-based metric

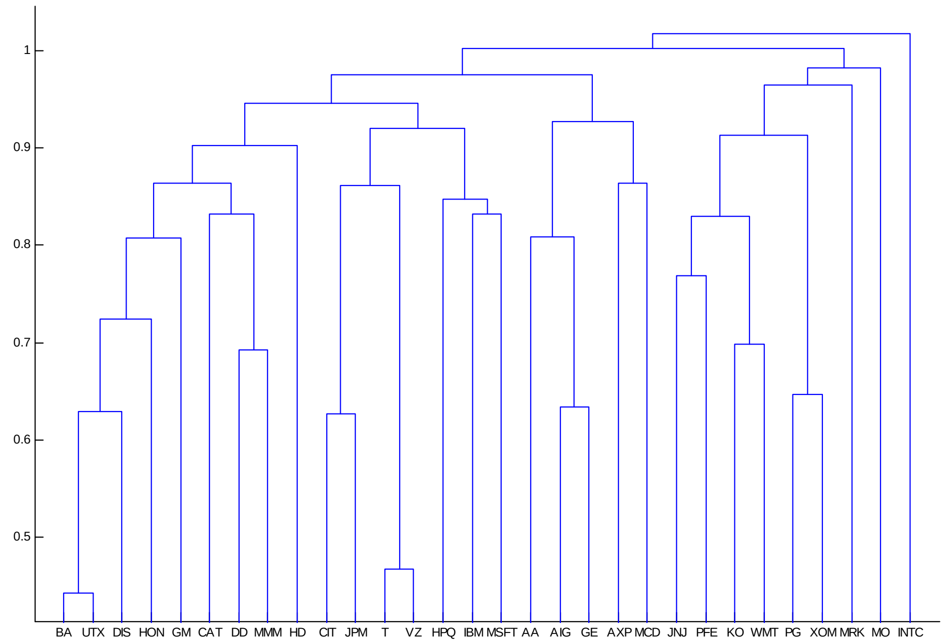

Figure 5. Dendrogram of DJIA squared returns using the correlation-based metric

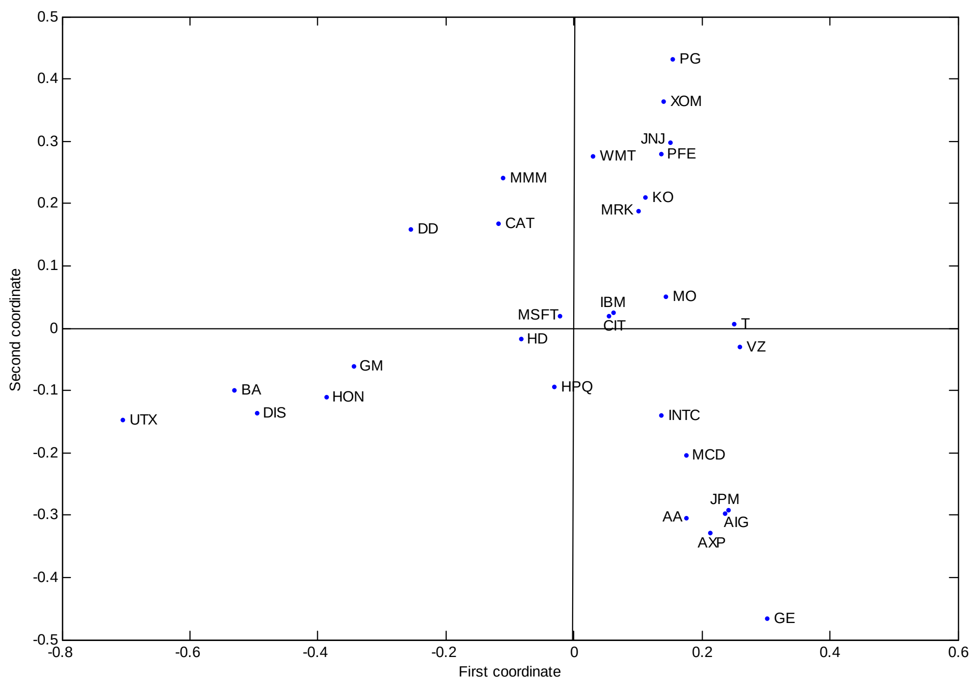

Figure 6. Multidimensional scaling of DJIA squared returns using the correlation-based metric

6. Conclusions

In this paper, we introduced a asymmetric-volatility and spectral based metric for clustering financial time series. Our methodological contribution consists of adding the asymmetry to the comparison factors and providing a weighted measure that takes into account both the spectral characteristics and the asymmetry.

Using the information about the simple TARCH model estimates and the log normalized periodogram ordinates of the squared returns, we investigated the similarities among the stock of the Dow Jones Industrial Average (DJIA) index. From empirical study, we found homogenous clusters of stocks with respect to the economic sectors. The financial companies and the technology companies are close together, the basic materials companies and services companies are in a distinct cluster, and most consumer and industrial Good companies are at the same location and close to the financial-technology cluster.

For reference, we also investigated the linkage between these stocks using cross-correlation methods. The two procedures give different, but not completely distinct results, which is very reassuring. The introduction of our technique allows for a more reliable differentiation between the series.

Acknowledgment: This research was supported by a grant from the Fundação para a Ciência e a Tecnologia (FEDER/POCI 2010). The authors gratefully acknowledge the helpful comments and suggestions of three anonymous referees on an earlier version of this paper.

REFERENCES

1. Ang, A. and Chen, J. (2002). "Asymmetric correlations of equity portfolios", *Journal of Financial Economics*, 63, 443-494.
2. Bekaert, G. and Wu, G. (2000). "Asymmetric volatility and risk in equity markets", *Review of Financial Studies*, 13, 1-42.
3. Bollerselev, T. Chou, R. and Kroner, K. (1992). "ARCH modeling in Finance", *Journal of Econometrics*, 52, 5-59.
4. Bonanno G., Lillo F., and Mantegna, R. (2001). "High-frequency cross-correlation in a set of stocks", *Quantitative Finance*, 1, 96-104.
5. Bonanno, G., Caldarelli, G., Lillo, F., Miccieché, S., Vandewalle N. and Mantegna, R. (2004). "Networks of equities in financial markets", *European Physical Journal B*, 36, 363-371.
6. Caiado, J., Crato, N. and Peña, D. (2006). "A periodogram-based metric for time series classification", *Computational Statistics & Data Analysis*, 50, 2668-2684.
7. Capiello, L., Engle, R. and Sheppard, K. (2006). "Asymmetric dynamics in the correlation of global equity and bond returns", *Journal of Financial Econometrics*, 4, 537-572.
8. Christie, A. (1982). "The stochastic behaviour of common stock variances: value, leverage and interest rate effects", *Journal of Financial Economics*, 10, 407-432.
9. Engle, R. and Ng, V. (1993). "Measuring and testing the impact of news on volatility", *Journal of Finance*, 48, 1022-1082.

10. Engle, R. and Kroner. K. (1995). "Multivariate simultaneous generalized ARCH", *Econometric Theory*, 11, 122-150.
11. Everitt, B., Landau, S. and Leese, M. (2001). *Cluster Analysis*, 4th ed., Edward Arnold, London.
12. Glosten, L. Jagannathan, R. and Runkle, D. (1993). "On the relation between the expected value and the volatility of the nominal excess return on stocks", *The Journal of Finance*, 48, 1779-1801.
13. Johnson, R. and Wichern, D. (2002). *Applied Multivariate Statistical Analysis*. 5th Ed., Prentice-Hall.
14. Kroner, K. and Ng, V. (1998). "Modeling asymmetric comovements of asset returns", *Review of Financial Studies*, 11, 817-844.
15. Mantegna, R. N. (1999). "Hierarchical structure in financial markets", *The European Physical Journal B* 11, 193-197.
16. McLeod, A. and Li, W. (1983). "Diagnostic checking ARMA time series models using squared-residual autocorrelations", *Journal of Time Series Analysis*, 4, 269-273.
17. Morrison, D. (2005). *Multivariate Statistical Methods*, 4th ed., Duxbury, Brooks/Cole Thomson Learning, Belmont.
18. Nelson, D. (1991). "Conditional heteroskedasticity in asset returns: a new approach", *Econometrica*, 59, 347-370.
19. Schwert, G. (1989). "Why does stock market volatility change over time?", *Journal of Finance*, 44, 1115-1153.
20. Tsay, R. (2005), *Analysis of Financial Time Series*, 2nd ed., Wiley, New Jersey.
21. Yu, C. and Wu, C. (2001). "Economic sources of asymmetric cross-correlation among stock returns", *International Review of Economics & Finance*, 10, 19-40.
22. Zakoian, J. (1994). "Threshold heteroskedasticity models", *Journal of Economic Dynamics and Control*, 18, 931-944.
23. Zivot, E. and Wang, J. (2003). *Modeling Financial Time Series with S-Plus*. Springer-Verlag, New York.