Identifying common dynamic features in stock returns

Caiado, Jorge and Crato, Nuno

April 2009

Identifying common dynamic features in stock returns

Jorge Caiado and Nuno Crato

CEMAPRE, Instituto Superior de Economia e Gestão,

Technical University of Lisbon,

Rua do Quelhas 6, 1200-781 Lisboa, Portugal.

Tel. +351 213 922715 Fax +351 213 925846

Email (corresponding author): jcaiado@iseg.utl.pt

Abstract

This paper proposes volatility and spectral based methods for cluster analysis of stock returns. Using the information about both the estimated parameters in the threshold GARCH (or TGARCH) equation and the periodogram of the squared returns, 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.

Keywords: Asymmetric effects; Cluster analysis; DJIA stock returns; Periodogram; Threshold GARCH 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 market analysis, the identification of similarities in index returns and volatilities can be useful for grouping countries. Finally, the existence of asymmetric cross-correlations and dependences in asset returns can be of interest for financial research.

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. These include the multivariate generalized autoregressive conditionally heteroskedasticity (GARCH) models of Engle and Kroner (1995) and Kroner and Ng (1998). To avoid these problems, various types of multivariate statistical techniques have been used for analyzing the structure of the asset returns. A first technique is the principal component analysis (PCA), which is concerned with the covariance structure of asset returns and can be used in dimension reduction (Tsay, 2005). A second technique is the factor model for asset returns that uses multiple time series to describe the common factors of returns (see, e.g., Zivot and Wang, 2003, for further discussion). A third technique 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 economic and 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 has two problems. Firstly, it does not take into account the stochastic volatility dependence of the processes – in fact, two processes may be highly correlated and have very different internal stochastic dynamics. Secondly, it cannot be used directly for comparison and grouping stocks with unequal sample sizes – this is a common problem of most existing nonparametric-based methods, as discussed, for instance, in Caiado, Crato and Peña (2009).

In this paper, we introduce a distance measure between the threshold GARCH model 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. Finally, 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.

The remaining sections are organized as follows. Section 2 provides volatility and spectral based distances for clustering asset returns. Section 3 describes the data and explores the univariate statistics. Section 4 presents the empirical findings on the cluster analysis. Section 5 covers the multidimensional scaling results. Section 6 summarizes and concludes.

2 Volatility and spectral based distances

Many time-varying volatility models have been proposed to capture the so-called "asymmetric volatility" effect (for a review, see the surveys by Bollerselev, Chou and Kroner, 1992, Kroner and Ng, 1998 and Bekaert and Wu, 2000), where volatility tends to be higher after a negative return shock than a positive shock of the same magnitude.

A univariate volatility model commonly used to allow for asymmetric shocks to volatility is the threshold GARCH (or TGARCH) model (see Glosten, Jagannathan and Runkle, 1993 and Zakoian, 1994). The simple TGARCH(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. The volatility may either diminish ($\gamma < 0$), rise ($\gamma > 0$), or not be affected ($\gamma \neq 0$) by negative shocks or "bad news" ($\varepsilon_{t-1} < 0$). Good news have an impact of α while bad news have an impact of $\alpha + \gamma$. The persistence of shocks to volatility can be given by $\alpha + \beta + \gamma/2$.

Nelson (1991) also proposed 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.

In real applications, z_t is often assumed to follow a "fat-tailed" distribution, as it can be given by the Generalized Error Distribution (GED). The GED has probability density function

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

where v is the tail-thickness 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 fat-tailed distributed. When $v = 2$, $\{z_t\}$ is normally distributed. When $v > 2$, $\{z_t\}$ is thin-tailed distributed. For details, see, e.g., Tsay 2005, p. 108.

We now introduce a distance measure for clustering time series with similar volatility dynamics 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 TGARCH(1,1) model to both time series by the method of maximum likelihood assuming GED innovations. Let $T_x = (\widehat{\alpha}_x, \widehat{\beta}_x, \widehat{\gamma}_x, \widehat{v}_x)'$ and $T_y = (\widehat{\alpha}_y, \widehat{\beta}_y, \widehat{\gamma}_y, \widehat{v}_y)'$ be the vectors of the estimated ARCH, GARCH, leverage effect and tail-thickness parameters, with the estimated covariance matrices given by V_x and V_y, respectively.

A Mahalanobis-like distance between the dynamic features of the return series $r_{x,t}$ and $r_{y,t}$, called the TGARCH-based distance, can be defined by

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

where $\Omega = V_x + V_y$ is a weighting matrix. This way, the matrix Ω weights the parameters taking into account the uncertainty in their estimation. The distance (5) takes into account the information about the stochastic dynamic structure of the time series volatilities and allows for unequal length time series.

We can also use methods based on the periodogram ordinates or the autocorrelations lags of the squared returns. The spectrum of the squared return series provides useful information about the time series behavior in terms of the ARCH effects.

Let $P_x(\omega_j) = n^{-1}|\sum_{t=1}^{n} r_{t,x} e^{-it\omega_j}|^2$ 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 $n/2$). Let 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

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

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 $r_{x,t}^2$ and $r_{y,t}^2$, respectively.

Since the parametric features of the TGARCH model are not necessarily associated with all the periodogram ordinates, the parametric and nonparametric approaches can be combined to take into account both the volatility dynamics and the cyclical behavior of the return series, that is

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

where $\lambda_i, i = 1, 2$ are normalizing/weighting parameters. We have chosen to balance the contributions of each component. Each normalizing parameter has been set as the inverse of the sample standard deviation of the corresponding pairwise distances. This way, higher uncertainty in the estimates is translated with a smaller weight, and higher confidence in the estimates is translated with a larger weight. In practice, the researcher may try a range of parameters, looking for a specific combination that better groups the series under consideration. Further research will probably lead to better rules, but at this moment we believe that trying a range of parameters may be the best strategy to assess the robustness of the conclusions.

It is straightforward to show that the statistics (5) and (8) satisfy the following distance properties: (i) $d(x, y)$ is asymptotically zero for independent time series generated by the same data generating process (DGP); (ii) $d(x, y) \geq 0$ as all the quantities are nonnegative; and (iii) $d(x, y) = d(y, x)$, as all transformations are independent of the ordering. However, nothing guarantees the triangle inequality, which is the remaining defining property of a distance. This is not a problem for the clustering algorithms we have used (Gordon, 1996, p. 66-67, and Johnson and Wichern, 2007, p. 674).

3 Data

The data used in this article consists of time series 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 corporations), Boeing, Caterpillar and Honeywell (industrial goods), Walt Disney, Home Depot, and McDonalds (services), 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 corporations and basic materials corporations have a positive skewness coefficient. There are no significant autocorrelations up to order 20 in the returns for corporations Boeing, Caterpillar, El Dupont, Walt-Disney, General Electric, General Motors, Honeywell, IBM, JP Morgan Chase and McDonalds.

Table 3 presents the estimation results of TGARCH(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 Caterpillar, Walt Disney, General Electric and Merck, and the leverage-effect for Inter-Tel Inc. and 3M Co., which are not significant at conventional levels. The distribution of the innovation series is fat-tailed for all stocks. As expected, the estimated persistence $(\widehat{\alpha} + \widehat{\beta} + \widehat{\gamma}/2)$ for all the asymmetric models is 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 Ljung-Box test statistic shows evidence of no serial correlation in the squared residuals up to order 20 for all stocks except Caterpillar, McDonalds and Verizon. In terms of the mean equation, the Ljung-Box test statistic does not reject the null hypothesis of no serial correlation in the model residuals for all stocks except American Int. Group, Johnson & Johnson, Pfizer, United Technologies, Verizon and Exxon Mobile.

4 Cluster analysis

Cluster analysis of time series attempts to determine groups (or clusters) of objects in a multivariate data set. Let k be the number of objects (time series) under consideration. The most commonly used partition clustering method is based in hierarchical classifications of the objects. In hierarchical cluster analysis, we begin with each object being considered as a separate cluster (k clusters). In the second stage, the closest two groups are linked to form $k-1$ clusters. The process continues until the last stage, in which all the objects are in the same cluster (see Everitt, Landau and Leese, 2001 for further discussion).

The dendrogram is a graphical representation of the results of the hierarchical cluster analysis. Clusters are connected by arches in a tree-like plot. The height of each arch represents the distance between the two clusters being considered.

The dendrogram shows how clusters are formed at each stage of the procedure. At the bottom, each object (time series) is considered its own cluster. The objects continue to combine upwards. At the top, all objects are grouped into a single cluster. In general, it is difficult to decide where to cutoff the lines and consider the clusters. Choices are usually debatable.

For our analysis, we first used the TGARCH-based distance defined in (5). Figure 1 shows the corresponding dendrogram for the DJIA stock returns, obtained by the complete linkage method (see, e.g., Johnson and

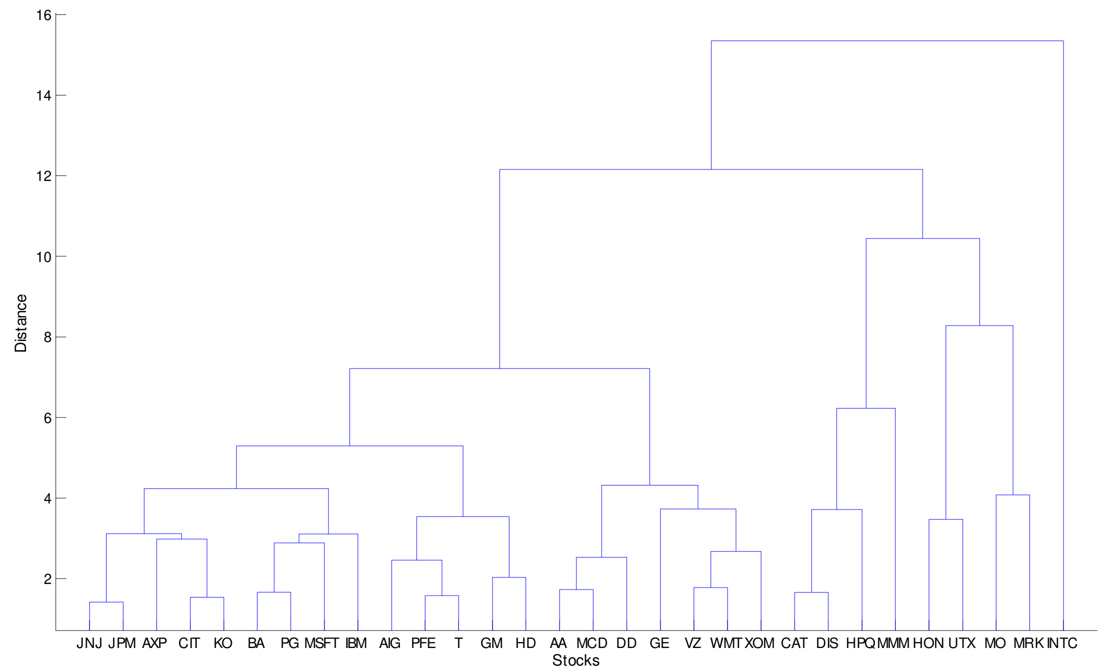

Figure 1: Complete linkage dendrogram for DJIA stocks using the Mahalanobis-TGARCH distance

Wichern, 2007).

As we want to use a sensible number of groups, this dendrogram suggests three to five clusters. We decided to consider five clusters. One is composed of most financial, consumer goods and healthcare corporations, some technology corporations (IBM, Microsoft and AT&T) and Home Depot and Boeing. The second is composed of basic materials and most services corporations and General Electric and Verizon. The third is composed of miscellaneous sector corporations (Caterpillar, Walt-Disney, Hewlett-Packard and 3M Co.). The fourth is composed of the industrial goods corporation Honeywell and the conglomerate corporation United Technologies. The fifth is composed of the consumer goods corporation Altria and the healthcare corporation Merck. The Inter-Tel corporation is not grouped.

Secondly, we used the spectral based distance defined in (6). Figure 2 shows the corresponding complete linkage dendrogram. We found three groups of corporations. One group is composed of basic materials (Alcoa, El Dupont and Exxon Mobile), communications (AT&T and Verizon), healthcare (Johnson & Johnson and Pfizer), financial (AIG and Caterpillar), and services (McDonalds and Walt-Mart Stores) corporations. The

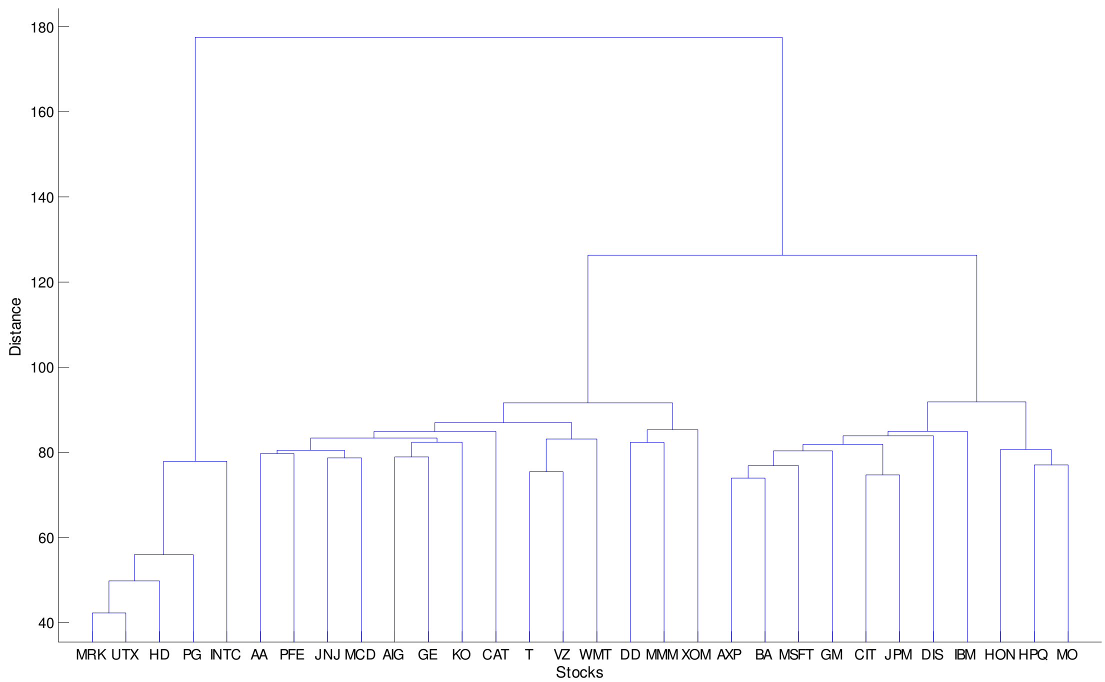

Figure 2: Complete linkage dendrogram for DJIA stocks using the LNP-based distance

second group is composed of technology (IBM, Microsoft and Hewlett-Packard), financial (American Express and JP Morgan Chase), industrial goods (Boeing, Citigroup and Honeywell), and consumer goods (Altria and General Motors) corporations. The third group is composed of miscellaneous sector corporations (Merck, United Technologies, Home Depot, Procter & Gamble and Inter-Tel).

Thirdly, we used the combined TGARCH-LNP based distance defined in (8). Figure 3 shows the corresponding complete linkage dendrogram. From the dendrogram, we can see three groups of corporations. One is formed by technology (IBM, Microsoft and Hewlett-Packard), financial (American Express, JP Morgan Chase and Caterpillar) and industrial goods (Boeing and Citigroup) corporations. The second group is formed by basic materials (Alcoa, El Dupont and Exxon Mobile), communications (AT&T and Verizon), healthcare (Johnson & Johnson and Pfizer) and services (McDonalds and Walt-Mart Stores) corporations. The third group is formed by consumer goods corporations (Altria and Procter & Gamble) and by a miscellaneous sector group (Home Depot, United Technologies, Honeywell and Merck). The corporations 3M Co. and Inter-Tel are not grouped.

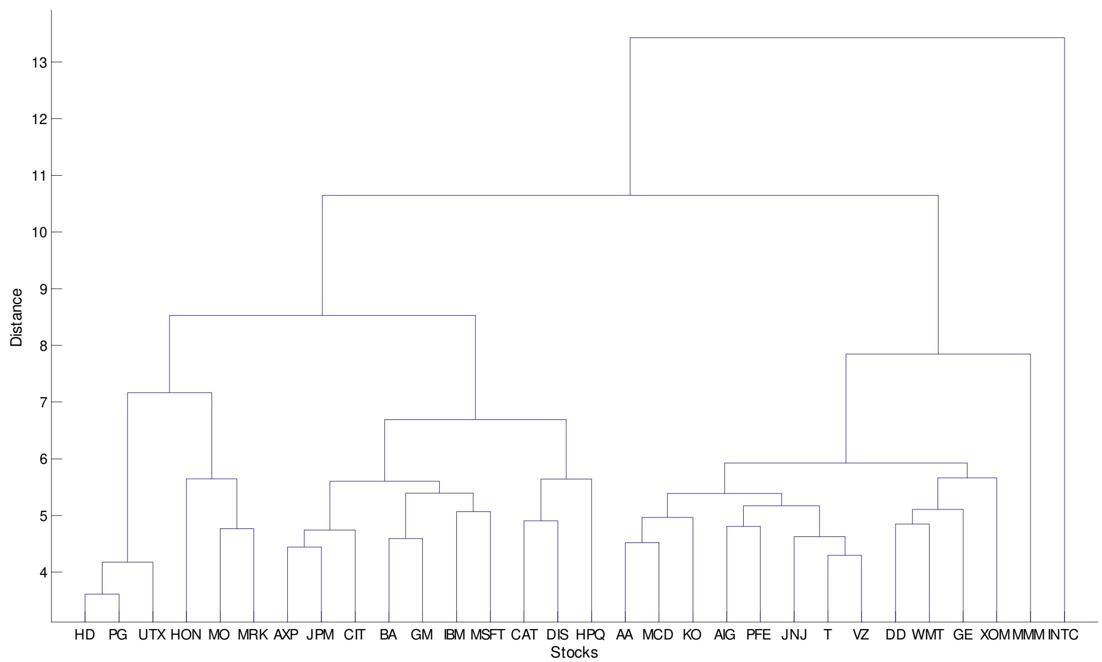

Figure 3: Complete linkage dendrogram for DJIA stocks using the combined LNP-TGARCH distance

5 Multidimensional scaling

Multidimensional scaling is a multivariate statistical method that uses the information about the similarities (or dissimilarities) between the objects (in our case, time series) to construct a configuration of k points in the p-dimensional space. See, for instance, Everitt and Dunn (2001) and Johnson and Wichern (2007).

Let D be the observed $k \times k$ matrix of Euclidean distances. By multidimensional scaling, the matrix D yields a $k \times p$ configuration matrix T. The rows of T are the coordinates of the k points in a p-dimensional representation of the observed dissimilarities ($p < k$). The p-dimensional representation that best approximates the observed dissimilarity matrix is given by the p eigenvectors of TT' corresponding to the p largest eigenvalues.

When the observed dissimilarity matrix D is not Euclidean, the matrix TT' is not positive semi-definite. In such cases some of the eigenvalues of TT' will be negative. If, however, the sum of the positive eigenvalues of TT' is approximately equal to the sum of all the eigenvalues and the magnitude of the largest positive eigenvalues exceeds clearly that of the largest negative eigenvalue, the spatial configuration of the observed dissimilarity matrix may still be advisable (Sibson, 1979).

As in the previous section, we will discuss separately the results of the three considered methods: the TGARCH, the LNP, and the combined TGARCH-LNP.

Firstly, table 4 shows the eigenvalues resulting from TGARCH distances between stocks and the eigenvectors associated with the first two eigenvalues. Since D is non-Euclidean distance, some eigenvalues are negative. The first eigenvalue is equal to 54.0% of the sum of all the eigenvalues (583.5). The second eigenvalue is equal to 23.0% of the sum of all the eigenvalues. The sum of the first four positive eigenvalues (565.1) is almost equal to the sum of all the eigenvalues. The magnitude of the first two eigenvalues (315.1 and 134.1) exceed clearly the magnitude of the largest negative eigenvalue (-37.2). The resulting solution fulfills the trace and magnitude adequacy criterions of Sibson (1979).

The size criterions of Mardia, Kent and Bibby (1979) suggest using the eigenvectors associated with the first two eigenvalues to represent the distances among stocks. Figure 4 shows the two-dimensional scaling map of the derived coordinate values. This plot can also help to identify the clusters.

Looking at the first coordinate of the derived representation, we notice that all basic materials and services corporations and most financial, consumer goods, technology and healthcare corporations appear close together. The industrial goods corporations Honeywell and Boeing

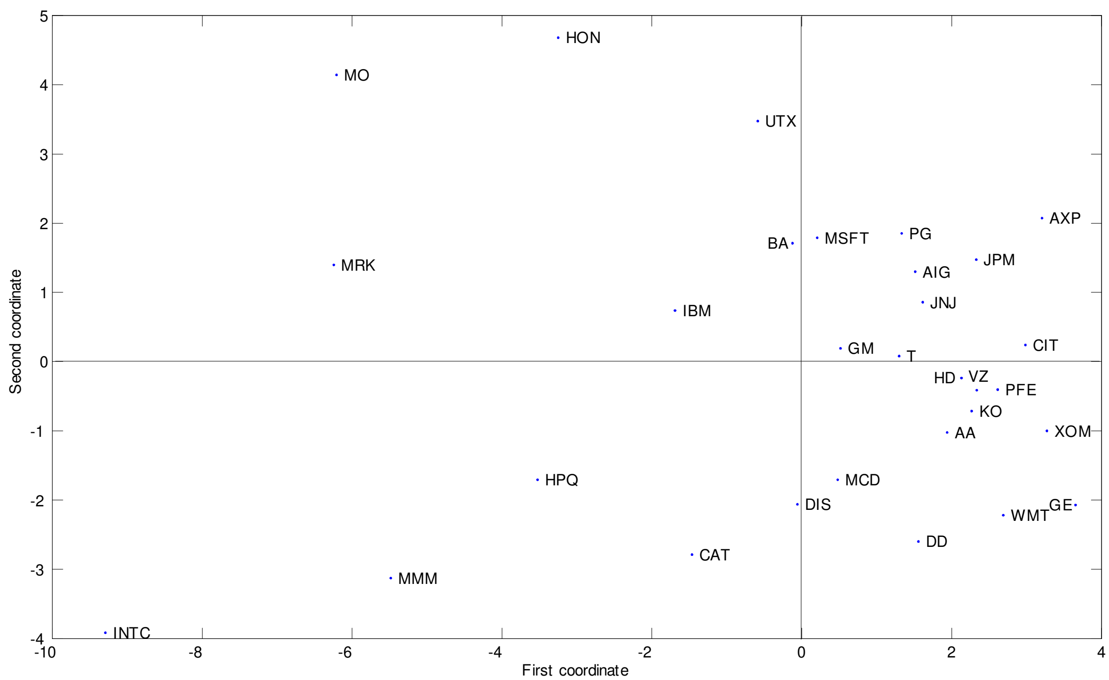

Figure 4: Two-dimensional scaling map of DJIA stocks using the Mahalanobis-TGARCH distance

are clearly separated from each other and from the remainder industrial goods corporations. Conglomerate corporations (3M and United Technologies) are in different locations. Again, Inter-Tel corporation is a clear outlier.

This first coordinate seems to translate the distribution tail behavior. Stocks with estimated tail-thickness parameter υ close to 1 (INTC, MMM, MRK and MO) are in the negative region of the first coordinate, while those with estimated υ close to 1.5 are clustered at the positive region. This means that the higher probability of having extreme shocks is a first major factor for distinguishing the stocks.

Looking at the second coordinate, we notice that basic materials and services corporations have negative eigenvalues and tend to cluster together, and that most financial, technology, consumer goods and healthcare corporations appear to form a distinct group. Again, the two conglomerate corporations are very clearly separated from each other.

This second coordinate seems to incorporate the magnitude of the asymmetric shocks to volatility played by the γ-coefficient. This means that the asymmetry is a second major factor for distinguishing the stocks.

Secondly, we consider the LNP method. Figure 5 shows the corresponding scaling map of the DJIA stocks. The map tends to group the basic materials, the communications, and most healthcare, financial and services corporations in a distinct cluster and most technology, industrial goods and consumer goods corporations in another distinct cluster.

To better interpret the two principal coordinates of the LNP method, we have computed the smoothed log normalized periodograms for each of the 30 DJIA squared return series. Figure 6 shows the corresponding plots.

The spectral function estimates are very diverse and the dissimilarities arise in the whole range of coordinates. The interpretation is very difficult. We notice that the first coordinate reveals a separate group at the left in which most corporations have an atypical spectral shape (PG, MRK, UTX, and HD). For these corporations, the spectra does not display a slowly decreasing long-term power, i.e., do not decrease regularly from the low to the higher frequencies. The second coordinate is even harder to interpret. We only highlight a clear separation of the communications corporationss (VZ and T) from the others.

Thirdly, the scaling solution for combined TGARCH-LNP distances is shown in Figure 7. The scaling map results are consistent with the dendrogram in Figure 3. The map suggests a separation of the stocks into three main clusters. The first is composed of basic materials, communications, and most healthcare and services corporations. The second

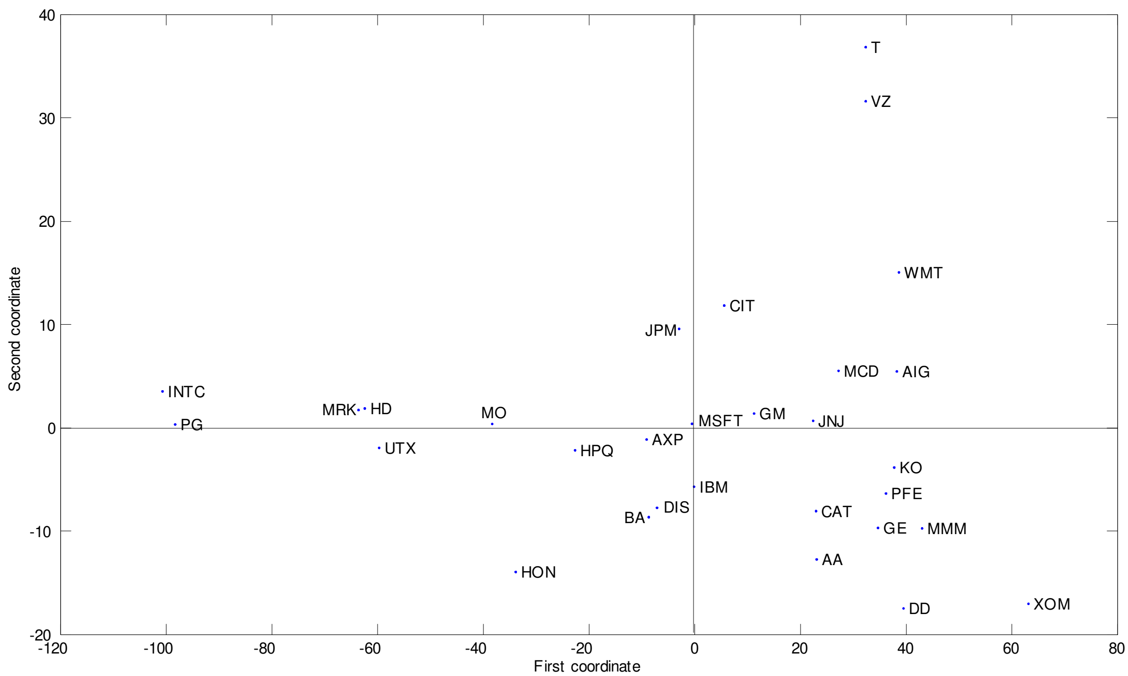

Figure 5: Two-dimensional scaling map of DJIA stocks using the LNP-based distance

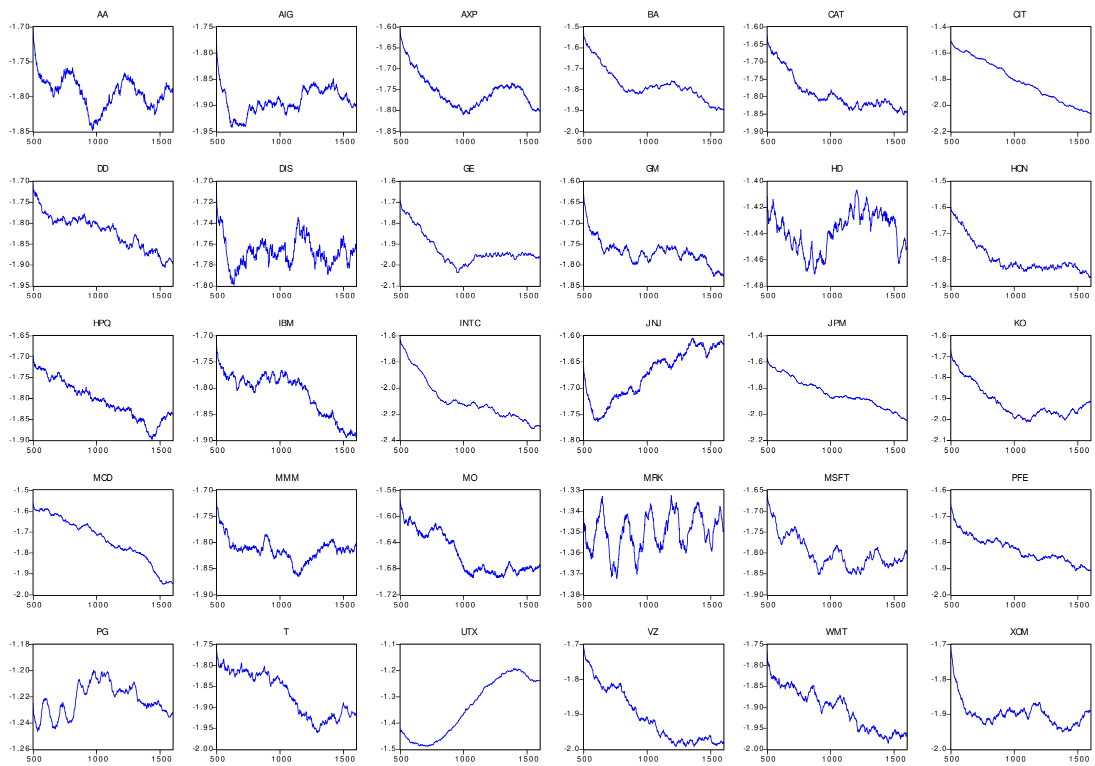

Figure 6: Smoothed log-normalized periodograms for DJIA squared return series

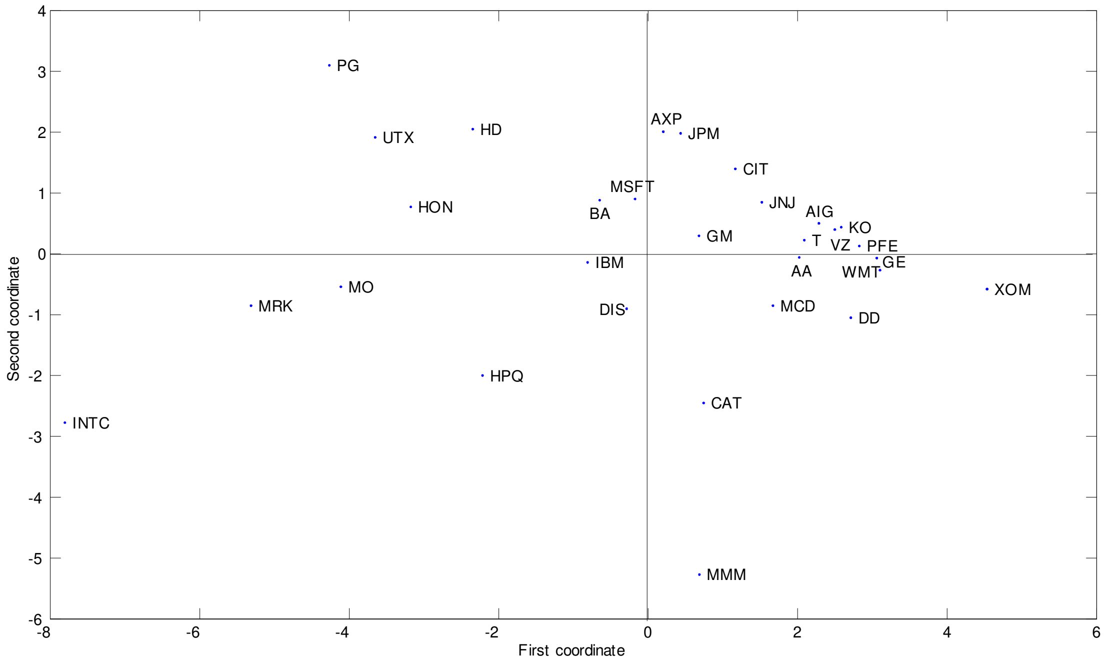

Figure 7: Two-dimensional scaling map of DJIA stocks using the combined LNP-TGARCH distance

is composed of most technology, financial and industrial good corporations. The third is composed of most consumer goods corporations and a miscellaneous sector corporations. Again, corporations with null and negative shocks on volatility (3M Co. and Inter-Tel) are in distinct locations and far from the other clusters.

The combined scaling map maintains the importance of the tail thickness for distinguishing the stocks (as we can see in the first map coordinate) and better clusters a central group.

6 Conclusions

In this paper, we introduced parametric and spectral-based distances for comparing and clustering multiple financial time series. Our methodological contribution consists essentially in adding the internal stochastic dynamic features to the comparison and in providing a combined distance that takes into account both the estimated model parameters and the spectral behavior of stocks' volatility.

We investigated the similarities among the stocks of the Dow Jones Industrial Average (DJIA) index. By using hierarchical clustering and multidimensional scaling techniques, we found that all considered methods (LNP, TGARCH, and combined TGARCH-LNP) are able to get meaningful corporate sector clusters. We found homogenous clusters of stocks with respect to the basic materials, services, healthcare, financial, communications and technology corporate sectors, and we found heterogeneous clusters of stocks with respect to the conglomerates, industrial goods, and consumer goods corporate sectors.

The TGARCH method tends to collect most financial, technology, consumer goods, and healthcare corporations into a cluster and basic materials and most services corporations into another one. The LNP method tends to group most technology and industrial good corporations into a cluster, and basic materials, communications and most healthcare corporations into another one. The combined TGARCH-LNP method tends to group most financial and technology corporations into a cluster, basic materials, communications and most healthcare corporations into another one, and most consumer goods into a third one.

In all cases, the thickness of the tail distribution plays an important discriminating role. This suggests that a higher probability of displaying extreme events seems to be an important factor in the classification of stocks. In the TGARCH method, the asymmetry parameter also plays an important role. This suggests that a different response to good and bad news is an important stock volatility discriminating factor.

The TGARCH and LNP methods led to somehow similar cluster solutions, which is very reassuring. The introduction of the combined TGARCH-LNP method allows for a potentially more reliable differentiation between the series, as it uses more information about the dynamic features of the stock returns and volatilities.

Acknowledgment: The authors are grateful to several anonymous referees for helpful comments and suggestions that vastly improved this article. They also thank the comments of participants in the COMPSTAT'2008 International Conference on Statistical Computing. This research was supported by a grant from the Fundação para a Ciência e a Tecnologia (FEDER/POCI 2010).

References

[1] Bekaert, G. and Wu, G. (2000). "Asymmetric volatility and risk in equity markets", *Review of Financial Studies*, 13, 1-42.

[2] Bollerselev, T. Chou, R. and Kroner, K. (1992). "ARCH modeling in Finance", *Journal of Econometrics*, 52, 5-59.

[3] Bonanno G., Lillo F., and Mantegna, R. (2001). "High-frequency

cross-correlation in a set of stocks", *Quantitative Finance*, 1, 96-104.

[4] Bonanno, G., Caldarelli, G., Lillo, F., Miccieché, S., Vandewalle N. and Mantegna, R. (2004). "Networks of equities in financial markets", *European Physical Journal B*, 38, 363-371.

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

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

[7] Engle, R. and Kroner. K. (1995). "Multivariate simultaneous generalized ARCH", *Econometric Theory*, 11, 122-150.

[8] Everitt, B. and Dunn, G. (2001). *Applied Multivariate Data Analysis*, 2th Ed., Hodder Arnold, London.

[9] Everitt, B., Landau, S. and Leese, M. (2001). *Cluster Analysis*, 4th Ed., Edward Arnold, London.

[10] 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.

[11] Gordon, A. (1996). "Hierarchical classification", in *Clustering and Classification*, P. Arabie, L. J. Hubert and G. De Soete (Eds.), World Scientific Publ., River Edge, NJ.

[12] Johnson, R. and Wichern, D. (2007). *Applied Multivariate Statistical Analysis*. 6th Ed., Prentice-Hall.

[13] Kroner, K. and Ng, V. (1998). “Modeling asymmetric comovements of asset returns”, *Review of Financial Studies*, 11, 817-844.

[14] Mantegna, R. N. (1999). "Hierarchical structure in financial markets", *The European Physical Journal B*, 11, 193-197.

[15] Mardia, K., Kent, J., and Bibby, J. (1979). *Multivariate Analysis*, Academic Press, London.

[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] Nelson, D. (1991). "Conditional heteroskedasticity in asset returns: a new approach", *Econometrica*, 59, 347-370.

[18] Sibson, R. (1979). "Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling", *Journal of the Royal Statistical Society B*, 41, 217-229.

[19] Tsay, R. (2005), *Analysis of Financial Time Series*, 2nd Ed., Wiley, New Jersey.

[20] Zakoian, J. (1994). "Threshold heteroskedasticity models", *Journal*

of Economic Dynamics and Control, 18, 931-944.
[21] Zivot, E. and Wang, J. (2003). *Modeling Financial Time Series with S-Plus*. Springer-Verlag, New York.

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	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 Technol.	UTX	Conglomerates
Hewlett-Packard	HPQ	Technology	Verizon Communic.	VZ	Technology
Int. Bus. Machines	IBM	Technology	Walt-Mart Stores	WMT	Services
Inter-tel Inc.	INTC	Technology	Exxon Mobile CP	XOM	Basic materials

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 TGARCH(1,1) models assuming GED innovations for DJIA stock returns

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

* (**) 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).

Table 4: Eigenvalues and eigenvectors resulting from TGARCH distances between DJIA stocks

Eigenvalues			First four eigenvectors					First four eigenvectors			
1-15	16-30	Stocks	1	2	3	4	Stocks	1	2	3	4
315.1	0.6	AA	1.94	-1.03	-0.95	-1.35	JNJ	1.61	0.86	-1.35	1.12
134.1	0.4	AIG	1.51	1.30	-0.67	0.05	JPM	2.33	1.47	-0.75	1.12
79.7	0.1	AXP	3.20	2.07	2.94	0.97	KO	2.27	-0.71	0.21	0.67
36.2	0.0	BA	-0.13	1.71	-0.08	-0.77	MCD	0.48	-1.71	0.26	-0.99
29.1	0.0	CAT	-1.46	-2.79	2.45	-0.48	MMM	-5.48	-3.13	2.27	-0.64
19.2	-0.1	CIT	2.98	0.24	0.64	0.72	MO	-6.21	4.14	-2.17	-1.94
12.5	-0.2	DD	1.56	-2.60	-0.04	-0.52	MRK	-6.24	1.40	0.77	-2.31
9.0	-0.3	DIS	-0.06	-2.06	3.05	-0.06	MSFT	0.21	1.79	-0.38	1.49
5.8	-0.8	GE	3.65	-2.07	-1.32	-0.15	PFE	2.61	-0.40	-0.71	-0.70
3.4	-1.0	GM	0.52	0.19	0.32	-0.30	PG	1.33	1.85	1.41	-0.76
2.4	-1.7	HD	2.13	-0.24	1.37	-0.04	T	1.30	0.08	-1.59	0.07
1.6	-4.0	HON	-3.25	4.68	-0.31	0.40	UTX	-0.59	3.48	1.51	1.04
1.3	-10.5	HPQ	-3.52	-1.71	2.15	0.38	VZ	2.34	-0.41	-1.85	0.29
1.1	-13.2	IBM	-1.69	0.74	0.36	2.07	WMT	2.69	-2.22	-1.23	-0.04
0.8	-37.2	INTC	-9.29	-3.92	-3.39	2.46	XOM	3.27	-1.00	-2.93	-1.78