A GARCH-based method for clustering of financial time series: International stock markets evidence

Jorge Caiado
School of Business Administration
Polytechnic Institute of Setúbal
Campus do IPS, Estefanilha
2914-503 Setúbal, Portugal
(e-mail: jcaiado@esce.ips.pt)

Nuno Crato
School of Economics and Business
Technical University of Lisbon
Rua do Quelhas, 6
1200-781 Lisboa, Portugal
(e-mail: ncrato@iseg.utl.pt)

July 28, 2007

Abstract

In this paper, we introduce a volatility-based method for clustering analysis of financial time series. Using the generalized autoregressive conditional heteroskedasticity (GARCH) models we estimate the distances between the stock return volatilities. The proposed method uses the volatility behavior of the time series and solves the problem of different lengths. As an illustrative example, we investigate the similarities among major international stock markets using daily return series with different sample sizes from 1966 to 2006. The data were divided into two sample periods: previous and subsquent to the terrorist attack on September 11, 2001. From cluster analysis in the period before 9-11, most European markets countries, United States and Canada appear close together, and most Asian/Pacific markets and the South/Middle American markets appear in a distinct cluster. After 9-11, the European stock markets have become more homogenous, and North American markets, Japan and Australia seem to come closer.

Keywords: Cluster analysis; GARCH model; International stock markets; Volatility.

1 Introduction

The general problem in clustering financial time series is the separation of a set of time series data into groups or clusters, with the property that series in the same group have a similar stochastic dependence structure and series in other groups are quite distinct. To perform cluster analysis of time series, we have to define a relevant measure of distance between the time series in a data set. The stochastic behavior of most financial time series renders the usual methodologies used to measure the distance between different stock returns inappropriate. Mantegna (1999), Bonanno, Lillo and Mantegna (2001), among others, used the Pearson correlation coefficient as similarity measure of a pair of stock returns. They computed a $k \times k$ matrix, where k is the number of stocks, with the $k(k-1)/2$ different pairs of correlation coefficients, and used the metric

$$d_{COR}(x, y) = \sqrt{2(1 - \widehat{\rho}_{xy})}, \tag{1}$$

where $\widehat{\rho}_{xy}$ is the correlation coefficient between the stock returns of the series x and y. Although this metric can be useful to ascertain the structure of stock returns movements, it has three important limitations: (i) it does not use the information about the autocorrelation structure of each stock return; (ii) it does not take into account the information about the return volatilities; and (iii) it cannot be used for comparison and grouping stocks with unequal sample sizes.

In this paper, we present a method for clustering analysis of financial time series without these drawbacks. First, we introduce a distance measure based on the generalized autoregressive conditional heteroskedasticity (GARCH) estimated models of the stock returns. We then investigate whether major international stock markets have similar volatility behavior. Previous studies have investigated the comovements of international equity returns by using mean correlations (see Longin and Solnik, 1995, Karolyi and Stulz, 1996, Mei and Ammer, 1996, Ramchmand and Susmel, 1998, Ball and Torous, 2000, and Morana and Beltratti, 2006), cointegration (see Arshanapalli and Doukas, 1993, Bessler and Yang, 2003, Syriopoulos, 2004, and Tahai, Rutledge and Karim, 2004), common factor analysis (see Engle and Susmel, 1993, and Hui, 2005), and other approaches. However, the problem of identifying similarities or dissimilarities in stock returns seems to be not enough explored in the empirical finance literature using cluster analysis.

The remainder of the paper is organized as follows. In Section 2, we introduce the parametric distance-based method for clustering of financial time series. In Section 3, we describe the data set used in this paper. In Section 4, we present the cluster analysis evidence for the empirical results. The final section summarizes the paper.

2 GARCH-feature based distance

We know that many of the recent finance time series theories are concerned with the conditional variance, or volatility, of a process. The volatility is a measure of the intensity of unpredictable changes in asset returns, so we can think of volatility as a random variable that follows a stochastic process. The task of any volatility model is to describe the historical pattern of volatility and possibly use this to forecast future volatility. Engle (1982) introduced the autoregressive conditional heteroskedasticity or ARCH(q) model assuming that the conditional variance depends on past volatility measured as a linear function of past squared values. The need of a long lag length q and the non-negativity conditions imposed in ARCH parameters led Bollerslev (1986) to propose a more parsimonious parameter structure model, the GARCH(p, q) model, defined by $\sigma_t^2 = c + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$. In most applications, the simple GARCH(1,1) model has been found to provide a good representation of a wide variety of volatility processes as discussed in Bollerslev, Chou and Kroner (1992).

We now introduce a parametric approach for clustering of financial time series using the information about the estimated GARCH parameters. Suppose we fit a GARCH(1,1) model to both return series r_x and r_y. Let $L_x = (\widehat{\alpha}_x, \widehat{\beta}_x)$ and $L_y = (\widehat{\alpha}_y, \widehat{\beta}_y)$ be the vectors of the estimated ARCH and GARCH parameters and V_x and V_y the estimated covariance matrices, respectively. Building upon the work of Caiado, Crato and Peña (2007), a measure of distance between the volatilities of the return series $r_{t,x}$ and $r_{t,y}$ can be defined by

$$d_{GARCH}(x, y) = (L_x - L_y)' V^{-1} (L_x - L_y), \tag{2}$$

where $V = V_x + V_y$. It is straightforward to show that this measure satisfies all the usual properties of a metric (except the triangle inequality): (i) $d(x, y) \geq 0$; (ii) $d(x, y) = 0$ if $x = y$; and (iii) $d(x, y) = d(y, x)$. The advantages of this measure over other distance-based methods are that it conveys all the stochastic structure of the conditional variance of a process and it solves the problem of comparison of time series with unequal length. We also should note that the proposed distance measure can be easily extended to larger GARCH models and to other type of volatility models.

3 Data description

We consider data of daily index returns for 27 international stock markets from Americas (Brazil, Argentina, Mexico, United States and Canada), from Asia/Pacific (India, Hong-Kong, Indonesia, Malaysia, Korea, Japan, Singapore, Taiwan, and Australia), from Europe (Netherlands, Austria, Belgium, France, Germany, United Kingdom, Spain, Italy, Sweden, Norway, and Switzerland), and from Middle East (Egypt and Israel), as reported in Table 1. These data were obtained from Yahoo Finance (http://finance.yahoo.com) and correspond to the adjusted close prices.

Table 2 contains the GARCH(1,1) estimates used to compute the volatility-based metric defined in (2). The sum of the ARCH and GARCH coefficients quantifies the shock persistence to volatility. A value of unity indicates a unit root in the conditional variance (see Engle and Bollerslev, 1986). The ARCH test is the Lagranger multiplier test for ARCH effects in the residuals (see Engle, 1982). The Q (Q^2) is the Ljung-Box test statistic for serial correlation in the residuals (squared residuals). In the GARCH models, all estimated coefficients are significant at conventional levels and have the appropriate signs. The shock persistences to volatility are close to one for all the markets. For Malaysia and Egypt, the summation of ARCH and GARCH estimates is slightly higher than 1. The diagnostic tests show that the models for all the stock markets perform well

Table 1: Daily indices of international stock markets

Stock market	Country	Period	Sample size
New York Stock Exchange	United States (US)	1966 - 2006	10259
TXS Venture Exchange	Canada (CAN)	2000 - 2006	1716
Sao Paolo Stock Exchange	Brazil (BRA)	1993 - 2006	3329
Buenos Aires Stock Exchange	Argentina (ARG)	1996 - 2006	2473
Mexico Stock Exchange	Mexico (MEX)	1991 - 2006	3721
Bombay Stock Exchange	India (IND)	1997 - 2006	2293
Hong Kong Stock Exchange	Hong-Kong (HK)	1987 - 2006	4890
Jakarta Stock Exchange	Indonesia (IND)	1997 - 2006	2233
Kuala Lumpur Stock Exchange	Malaysia (MAL)	1993 - 2006	3165
Korea Stock Exchange	Korea (KOR)	1997 - 2006	2277
Japan Stock Exchange	Japan (JAP)	1984 - 2006	5602
Singapore Stock Exchange	Singapore (SING)	1987 - 2006	4692
Taiwan Stock Exchange	Taiwan (TAI)	1997 - 2006	2277
Australian Stock Exchange	Australia (AUST)	1984 - 2006	5607
Amsterdam Stock Exchange	Netherlands (NET)	1992 - 2006	3557
Vienna Stock Exchange	Austria (AUS)	1992 - 2006	3437
Brussels Stock Exchange	Belgium (BEL)	1991 - 2006	3899
Paris Stock Exchange	France (FRA)	1990 - 2006	4185
Xetra Stock Exchange	Germany (GER)	1990 - 2006	4000
London Stock Exchange	United Kingdom (UK)	1984 - 2006	5687
Madrid Stock Exchange	Spain (SPA)	1993 - 2006	3321
Milan Stock Exchange	Italy (ITA)	2000 - 2006	1752
Stockholm Stock Exchange	Sweden (SWE)	2001 - 2006	1452
Oslo Stock Exchange	Norway (NOR)	2001 - 2006	1429
Swiss Stock Exchange	Switzerland (SWI)	1990 - 2006	4001
Egypt Stock Exchange	Egypt (EGY)	1997 - 2006	1815
Tel Aviv Stock Exchange	Israel (ISR)	1997 - 2006	1853

in terms of the variance equation except Brasil, United Kingdom, Hong-Kong, and Mexico, which show evidence of ARCH effects in the fitted residuals.

4 Cluster analysis

To investigate the affinity between the major international stock markets, we perform a cluster analysis of the time series of daily stock-market indices using all available data for sample periods before and after the terrorist attack on Septemper 11, 2001. For each data set, we compute a distance matrix with $k(k-1)/2$ different pairs using the GARCH-based method. Then, by using dendrogram and multimidensional scaling techniques (see for instance, Johnson and Wichern, 1992) based on the computed distances, we display clusters for the return series.

Table 2: Estimates for the international stock-market volatilities based on the GARCH(1,1) model

Market	ARCH	GARCH	Persistence	$Q(20)$	$Q^2(20)$	$LM(20)$
United States	0.08017	0.90451	0.98468	230.40*	16.86	16.84
Canada	0.05254	0.94309	0.99563	19.83	10.30	9.47
Brazil	0.11041	0.87385	0.98426	79.23*	32.56**	32.01**
Argentina	0.11909	0.85635	0.97544	44.47*	17.06	17.43
Mexico	0.11416	0.86893	0.98309	122.55*	39.34*	38.88*
India	0.11926	0.84775	0.96701	69.44*	15.99	15.72
Hong-Kong	0.13476	0.84615	0.98091	95.12*	170.64*	181.59*
Indonesia	0.13141	0.84919	0.98060	117.62*	18.37	17.73
Malaysia	0.11713	0.88651	1.00364	130.43*	21.39	21.92
Korea	0.07183	0.92705	0.99888	32.91**	8.42	7.94
Japan	0.12237	0.87537	0.99774	32.57**	15.80	15.85
Singapore	0.15544	0.80763	0.96307	104.46*	3.46	3.50
Taiwan	0.07212	0.92181	0.99393	22.86	28.96	27.91
Australia	0.22299	0.69768	0.92067	91.10*	7.79	7.80
Netherlands	0.09020	0.90293	0.99313	31.95**	28.36	29.59
Austria	0.09491	0.86437	0.95928	72.41*	16.23	16.55
Belgium	0.10615	0.86154	0.96769	72.74*	4.30	4.24
France	0.07682	0.90647	0.98329	24.58	15.40	15.69
Germany	0.07827	0.90359	0.98186	27.53	3.10	3.02
United Kingdom	0.09030	0.89146	0.98176	35.20**	58.11*	56.61*
Spain	0.09379	0.89372	0.98751	30.24	14.16	13.35
Italy	0.08154	0.91051	0.99205	17.83	25.34	24.59
Sweden	0.09499	0.89237	0.98736	18.39	18.24	18.37
Norway	0.12810	0.80014	0.92824	29.03	19.02	19.09
Switzerland	0.12178	0.83409	0.95587	28.78	4.19	4.17
Egypt	0.18812	0.85540	1.03352	101.34*	10.84	10.37
Israel	0.09684	0.81474	0.91158	25.23	16.57	16.19

* (**) Significant at the 1% (5%) level.

4.1 Before September 11, 2001

Figure 1 presents the map of distances across international stock markets using the 2-dimensional GARCH scaling and the dendrogram by complete linkage algorithm from which the clusters of markets can be identified. We found that all the markets are nearly at the same first coordinate except Australia, United States and Canada. Looking at the second coordinate, we appear to have the major European markets close together, the South/Middle American markets are at the same position, and some Asian/Pacific markets are at the same location. From the dendrogram, we can split the indices returns into three distinct clusters: Cluster 1 = (FRA, ITA, AUS, GER, NET, KOR, US, BEL, SPA, UK, CAN); Cluster 2 = (IND, SWI, TAI, ISR, HK, INDO, ARG, SING, BRA, JAP, MAL, MEX); and Cluster 3 = (AUST, EGY). Cluster 1 includes eight of the major European markets (France, Germany, Italy, United Kingdom, Netherlands, Spain, Austria and Belgium), the North American countries (United States and Canada) and Korea. Cluster 2 includes the South/Middle American markets (Brazil, Mexico, and Argentina), seven of the major Asian/Pacific markets (Japan, Taiwan, Malaysia, Hong-Kong, India, Indonesia, and Singapore), Switzerland and Israel. Cluster 3 grouped the outliers Australia and Egypt.

4.2 After September 11, 2001

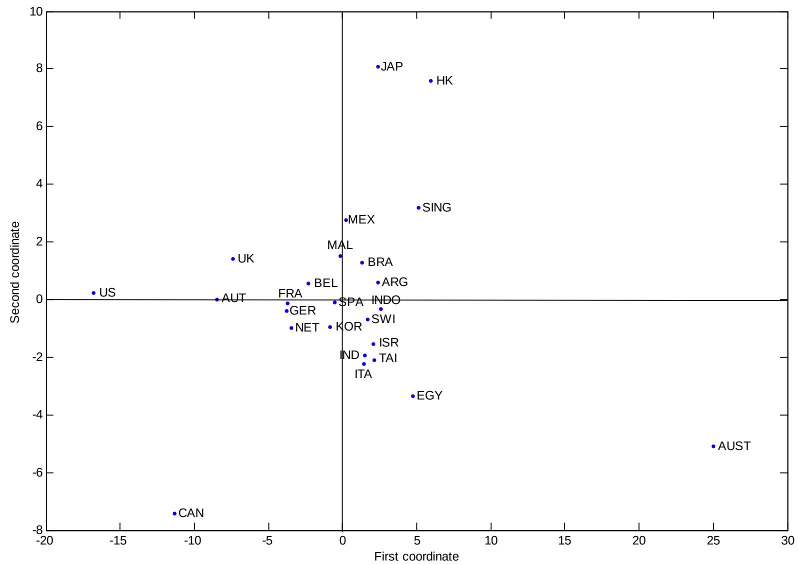
Figure 2 shows the distances across stock markets in the sample period from September 11, 2001 to 2006. We appear to have most developed countries United States, Canada, Australia, Germany and Japan close to each other, and close to European countries United Kingdom, France, Spain, Netherlands and Italy. Looking at the dendrogram, we found three very reasonably clusters: Cluster 1 includes eight European countries (Germany, France, Spain, Netherlands, United Kingdom, Switzerland, Belgium and Sweden), Japan, Singapore, Korea, Israel and Argentina; Cluster 2 includes United States, Canada, Australia, Italy, Taiwan, Hong-Kong, Egypt and Brazil; and Cluster 3 includes Austria, Norway, Malaysia, India, Indonesia and Mexico.

5 Conclusions

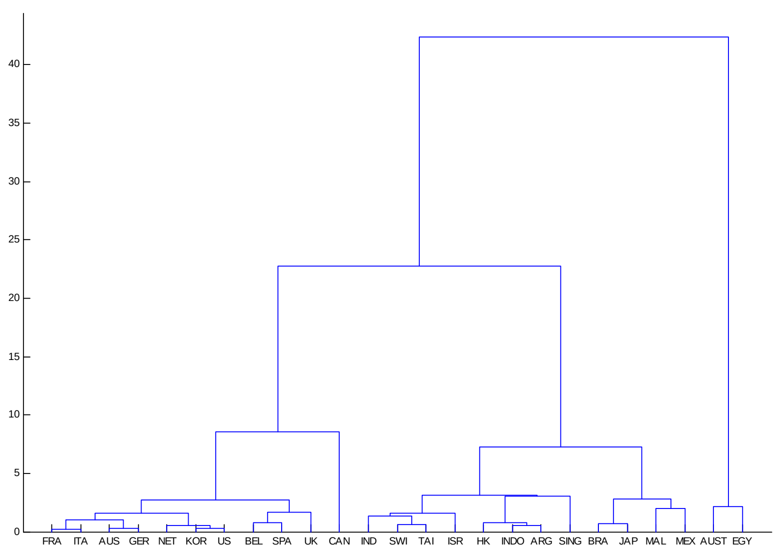
In this paper, we introduced a volatility-based method for comparison of financial time series, and we investigated the similarities among major international and stock-markets returns. The proposed method takes into account the stochastic volatility dependence of the processes and solves the problem of classification of time series with unequal length.

We performed a cluster analysis for daily stock indices returns with unequal sample sizes from 1966 to 2006. In our empirical study, we found that the persistence estimates are very similar for all stock markets except Australia, which makes it hard to identify dissimilarities among the stock market volatilities.

Figure 1: Distances across stock markets for the period before 9-11

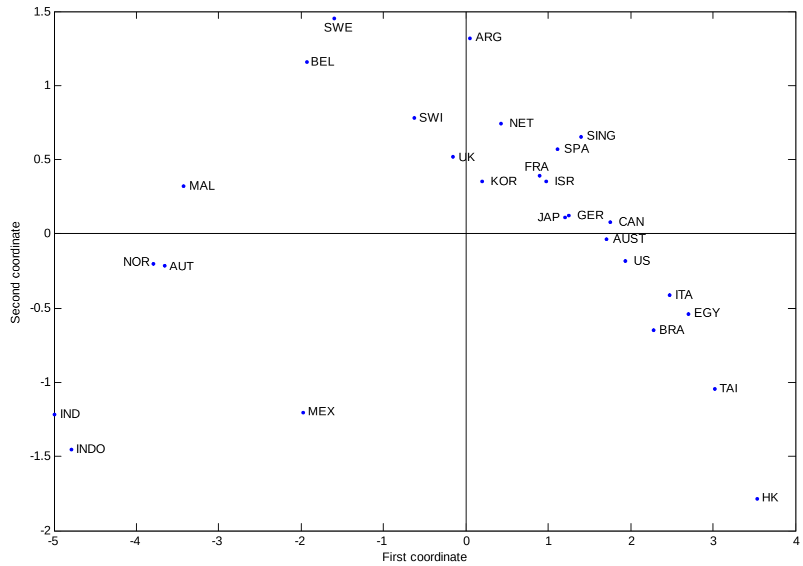


(a) Principal coordinates analysis

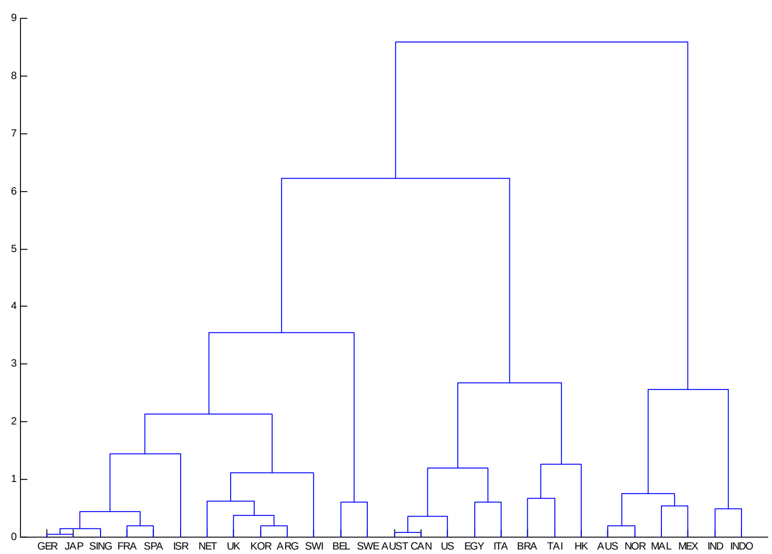


(b) Dendrogram by complete linkage

Figure 2: Distances across stock markets for the period after 9-11



(a) Principal coordinates analysis



(b) Dendrogram by complete linkage

However, using the GARCH-feature based method for the period before 11 September 2001, we found three distinct clusters. One cluster is formed by most European countries, United States, Canada and Korea. The second is formed by South/Middle American markets (Brazil, Argentina, and Mexico), the major Asian/Pacific markets (Japan, Taiwan, Hong-Kong, India, Malaysia, Indonesia, and Singapore), Israel and Switzerland. The third is formed by Australia and Egypt. The results are slightly different in the sample period after the terrorist attacks. The European countries seem to become more homogenous after 9-11, in part due to the euro area markets integration, and the United States, Canada, Australia and Japan markets tend to cluster together.

Acknowledgment: This research was supported by a grant from the Fundação para a Ciência e Tecnologia (FEDER/POCI 2010).

References

[1] Arshanapalli, B. and Doukas, J. (1993). "International stock market linkages: Evidence from the pre and post-October 1987 period", *Journal of Banking & Finance*, 17, 193-208.

[2] Ball, C. and Torous, W. (2000). "Stochastic correlation across international stock markets", *Journal of Empirical Finance*, 7, 373-388.

[3] Bessler, D. and Yang, J. (2003). "The structure of interdependence in international stock markets", *Journal of International Money and Finance*, 22, 261-287.

[4] Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

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

[6] Bonanno G., Lillo F., and Mantegna, R. N. (2001). "High-frequency cross-correlation in a set of stocks", *Quantitative Finance*, 1, 96-104.

[7] Caiado, J., Crato, N. and Peña, D. (2007). "Comparison of time series with unequal lengths", manuscript.

[8] Engle, R. (1982). "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, 50, 987-1008.

[9] Engle, R. and Bollerslev, T. (1986). "Modelling the persistence of conditional variances", *Econometric Reviews*, 5, 1-50.

[10] Hui, T. (2005). "Portfolio diversification: a factor analysis approach", *Applied Financial Economics*, 15, 821-834.

[11] Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. 3rd Ed., Englewood Cliffs, Prentice-Hall.

[12] Karolyi, G. and Stulz, R. (1996). "Why do markets move together? An investigation of U.S.-Japan return comovements", *The Journal of Finance*, 51, 951-986.

[13] Longin, F. and Solnik, B. (1995). "Is the correlation in international equity returns constant: 1960-1990?", *Journal of International Money and Finance*, 14, 3-26.

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

[15] Mei, J. and Ammer, J. (1996). "Measuring international economic linkages with stock market data", *The Journal of Finance*, 51, 1743-1763.

[16] Morana, C. and Beltratti, A. (2006). "Comovements in international stock markets", *Journal of International Financial Markets, Institutions and Money*, in press.

[17] Ramchmand, L. and Susmel, R. (1998). "Volatility and cross correlation across major stock markets", *Journal of Empirical Finance*, 5, 397-416.

[18] Syriopoulos, T. (2004). "International portfolio diversification to Central European stock markets", *Applied Financial Economics*, 14, 1253-1268.

[19] Tahai, A., Rutledge, R. and Karim, K. (2004). "An examination of financial integration for the group of seven (G7) industrialized countries using an I(2) cointegration model", *Applied Financial Economics*, 14, 327-335.