GARCH model-based fuzzy clustering of Asian and Oceania stock markets

Seyede Nafiseh Alemohammad
Shahed University, Tehran, Iran
Email: n.alemohammad@shahed.ac.ir

Abstract
This paper applies fuzzy clustering based on partitioning around medoids algorithm to partition financial time series. To measure the similarity or dissimilarity between employed time series, GARCH parametric approach is exerted. Based on the fuzzy and crisp silhouette criteria, some main Asian and Oceania stock markets are partitioned in two (the best choice for number of) clusters. The results indicate that the stock market of Thailand, Taiwan, Singapore, Australia, India, Philippines, South kore, Hong kong, Indonesia and China have similar trend with high degree membership.

Keywords: fuzzy clustering; medoids; time series; GARCH; silhouette criterion.

1. Introduction

In the last two decades, time series clustering has been grown with variant applications in many different fields, such as finance, economics, environmetrics, geology, see [6], [9] and [1]. Given a set of N data tuples, a clustering method creates k partitions where at least one object exists in each partition and $k < N$. The partitioning is crisp if each data point pertains to exactly one cluster, or fuzzy if each point may pertain to multiple clusters with different grades of membership. Financial time series relieve some particular specifications that must be considered when the objective is to partition them into homogenous clusters [6]. Some of these characteristics are as follows:

- They are usually non-stationary.
- There is almost no significant correlation between financial time series.
- The squared of the observations are strongly correlated.
- They almost depict leverage effect property i.e. the conditional variance of series reacts distinctly to positive and negative shocks with the same absolute values, [10].

These specifications of financial time series need to be taken in to account to evolve a proper distance measure for clustering. The Autoregressive Conditional Heteroscedasticity (ARCH) [8] and Generalized Autoregressive Conditional Heteroscedasticty (GARCH) [2] models are highly useful for modeling financial time series. In this paper, in order to cluster the main Asian and Oceania stock markets, a partitioning around medoids (PAM) algorithm is handled which the distance measure is based on the GARCH modeling, [6], [7] and [3].

The organization of this paper is as follows: the distance measure based on the GARCH model is illustrated in section 2. Fuzzy clustering around medoids is studied in section 3. Section 4 is dedicated to the main results of the implementation of proposed clustering model on some financial time series.

2. Time series models

2.1. Autoregressive moving average model

The time series $\{y_t\}$ is an Autoregressive moving average process of order p and q, ARMA(p,q), if it is stationary and satisfies

$$y_t - \varphi_1 y_{t-1} - \cdots - \varphi_p y_{t-p} = z_t + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q}, \tag{1}$$

Where $\{z_t\}$ is a white noise sequence with zero mean and constant variance σ^2. The left hand side of relation (1) is AR component of the model and the right-hand side give the MA component.

2.2. Generalized Auto regressive conditional heteroscedasticity model

The Generalized Autoregressive Conditional heteroscedasticity model of order p and q for time series $\{y_t\}$, GARCH(p,q), is introduced as [2]

$$y_t = z_t\sqrt{h_t} \tag{2}$$

where $\{z_t\}$ is a with noise with zero mean and variance σ^2 and h_t is the conditional variance of $\{y_t\}$, that is defined as

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i\, y_{t-i}^2 + \sum_{j=1}^{q} \beta_j\, h_{t-j}. \tag{3}$$

Sufficient conditions to warrent strictly positive h_t are $\omega > 0$, $0 \le \alpha_i, \beta_j < 1$, for $1 \le i \le p$ and $1 \le j \le q$.

Let $u_t = y_t^2 - h_t$ be the martingale difference (the expectation of u_t and the conditional expectation of it given the past information are zero). The squared of returns (y_t^2) can be exhibited as an ARMA(max(p,q),q) process as

$$y_t^2 = \omega + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_j) y_{t-i}^2 - \sum_{j=1}^{q} \beta_j u_{t-j} + u_t \tag{4}$$

where $\alpha_i = 0$ for $i > p$ and $\beta_i = 0$ for $i > q$. Under the stationarity and invertibility conditions of ARMA model, y_t^2 can be represented as an infinite AR model where θ_k is computed as

$$\theta_k = (\alpha_k + \beta_k) + \sum_{j=1}^{\min(k,q)} \beta_j\, \theta_{k-j} \tag{5}$$

and $\theta_0 = -1$.

3. Fuzzy clustering approach

In many real applications there is generally no explicit frontier between clusters, so that fuzzy clustering usually has better performance related to crisp clustering. In fuzzy clustering, the objects can pertain to multiple clusters and associated with each of the objects are membership values between zero and one which demonstrate the degree to which the item belongs to distinguished clusters, see [5].

3.1. GARCH model-based fuzzy clustering

Let $\mathrm{Y} = \{\mathrm{Y}_1, \mathrm{Y}_2, \dots, \mathrm{Y}_\mathrm{I}\}$ be a set of I financial time series, fitting the GARCH(p,q) model to each of them, the estimated AR coefficients of the squared time series are $\hat{\theta} = \{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_I\}$, where $\hat{\theta}_i$ is the parameter vector of i-th financial time series as $\hat{\theta}_i = \{\hat{\theta}_{1i}, \hat{\theta}_{2i}, \dots, \hat{\theta}_{D_i i}\}$ and D_i is the order of i-th AR process. Consider a subset of Y with cardinality C, $X = \{X_1, \dots, X_C\}$ with the AR coefficients vector $\hat{\eta} = \{\hat{\eta}_1, \hat{\eta}_2, \dots, \hat{\eta}_C\}$, where $\hat{\eta}_i = \{\hat{\eta}_{1i}, \hat{\eta}_{2i}, \dots, \hat{\eta}_{D_i i}\}$. The degree of membership of i-th time series in cluster c is derived as [6]

$$u_{ic} = \frac{1}{\sum_{z=1}^{C}\left[\frac{\sum_{r=1}^{D}(\hat{\theta}_{ri} - \hat{\eta}_{ri})^2}{\sum_{r=1}^{D}(\hat{\theta}_{ri} - \hat{\eta}_{rz})^2}\right]^{\frac{1}{m-1}}} \tag{6}$$

where $D = \max\{D_i, i = 1, \dots, I\}$ and m>1 is a weighting power that controls the fuzziness clustering [7].

3.2. Crisp silhouette and fuzzy silhouette criteria

One of the most popular criteria for decision making process in cluster analysis is the average silhouette with criterion. Consider i-th time series Y_i pertaining to cluster $c = 1,2, \dots, C$,with the membership degree u_{ic}, that u_{ic}>u_{id} foe every $e \in \{1,2, \dots, C\}$, e≠ c. Denote the mean distance of time series Y_i to all other series in cluster c by $\boldsymbol{a_{ci}}$ and the mean distance of this time series to all series pertaining to cluster e, $e \neq c$ by d_{ei}, also consider

$$b_{ci} = \min d_{ei}\,, \qquad e = 1,2, \dots . C, \quad c \neq e;$$

then the silhouette of time series Y_i is calculated as

$$S_i = \frac{b_{ci} - a_{ci}}{\max\{a_{ci}, b_{ci}\}},$$

and the crisp silhouette(CS) is defined the mean of the S_i over $i = 1,2, \dots, I$. A generalization of silhouette criterion, fuzzy silhouette is defined as

$$FS = \frac{\sum_{i=1}^{I} (\mu_{ci} - \mu_{ei})^{\alpha} S_i}{\sum_{i=1}^{I} (\mu_{ci} - \mu_{ei})^{\alpha}},$$

where μ_{ci} and μ_{ei} are the first and second largest elements of the i-column of the membership matrix and $\alpha \geq 0$. The good fuzzy partition are determined by large values of CS and FS [4].

4. Main results of empirical application

In this section, we exert the GARCH model-based fuzzy clustering to partition some main stock markets of Asian and Oceania for the sample period from 04/03/2013 to 01/01/2018 (about 1200 observations). Figure 1 plots the stock market returns of Tehran, Thailand, Taiwan, Singapre, Australia, India, New zealand, Philippines, South kore, Japan, Hong kong, Indonesia and China. The best value of the number of clusters by computing the crisp silhouette(CS) and fuzzy silhouette(FS) criteria is obtained two, see Table 1.

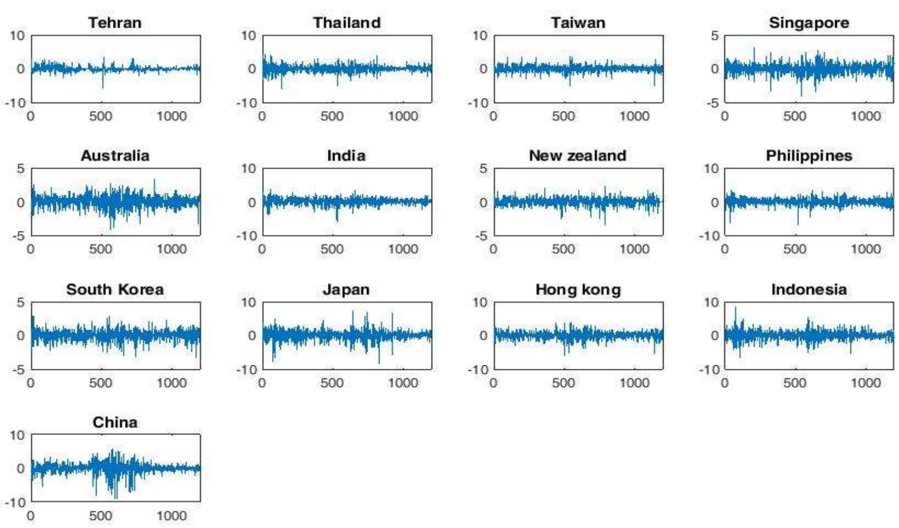

Figure 1. Stock market indexes

Table 1. Crisp and fuzzy silhouette measures

Number of clusters	CS	FS
2	**0.739**	**0.944**
3	0.470	0.291
4	0.384	0.122

The degrees of membership for each time series in clusters 1 and 2 are demonstrated in Table 2. The results of Table 2 show that except the stocks of Tehran, New zealand and Japan, the rest of the studied countries are in cluster 2 with strong membership degrees.

Table 2. The membership degree for stock markets

Countries	Cluster1	Cluster2	Countries	Cluster1	Cluster2
Tehran	0.666	0.333	Philippines	0.238	0.762
Thailand	0.012	0.988	SouthKorea	0.175	0.825
Taiwan	0.025	0.975	Japan	1	0
Singapore	0.031	0.969	Hong kong	0.067	0.933
Australia	0	1	Indonesia	0.007	0.993
India	0.067	0.933	China	0.054	0.946
New zealand	0.648	0.352			

References

[1] Aghabozorgi, S., Teh, Y.W. (2014). Stock market co-movement assessment using a three-phase clustering method. Expert Systems with Applications, 41(4), 1301-1314.

[2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticityJournal of Econometrics, 31, 307-327.

[3] Caiado, J ,Crato, N. (2007). CA GARCH-based method for clustering of financial time series: international stock markets evidence, in: C.H Skiadas (Ed.). Recent Advances in Stochastic Modeling and Data Analysis, World Scientific Publishing, Singapore, 542-551.

[4] Campello, R.J.G.B., Hruschka, E.R. (2006). A fuzzy extension of the silhouette width criterion for cluster analysis, Fuzzy Sets and Systems, 157, 2858-2875.

[5] D'urso, P., Maharaj, E.A, (2009). Autocorrelation-based fuzzy clustering of time series, Fuzzyn Sets and Systems, 160, 3565-3589.

[6] D'urso, P., Cappelli, C., Lallo, D.D, Massari, R., (2013). *Clustering of financial time series*, Physica A: Statistical Mechanics and its Applications, 392, 2114-2129.

[7] D'urso, P., Giovanni, L.D., Massari, R., (2016). *GARCH-based robust clustering of time series*, Fuzzyn Sets and Systems, 305, 1-28.

[8] Engle, R.F., (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, Econometrica, 50, 987-1007.

[9] Otranto, E., (2008). Clustering heteroscedastic time series by model-based procedures, Computational Statistics and Data Analysis, 52, 4685-4698.

[10] Tsay, R.S., (2005). Analysis of financial time series, 2nd ed., New York: John Wiley.