JEL Classification: C52, G11, G15, G32
Keywords: regime switching copula model, Spearman ratio, clustering stock indices

Grouping Stock Markets with Time-Varying Copula-GARCH Model*

Anna CZAPKIEWICZ—Faculty of Management, AGH University of Science and Technology, Cracow, Poland (gzrembie@cyf-kr.edu.pl), *corresponding author*

Paweł MAJDOSZ—STATMET, s.c., Cracow, Poland

Abstract

The aim of this work is to find the dynamics of interdependencies and similarities between European, American and Asian stock markets. The investigation covers daily returns of 36 market indices. In order to examine the dependencies between these data, the Markov regime switching copula model with two regimes is considered. For the dynamic clustering purposes, the time varying Spearman ratio obtained from the regime switching copula model is taken to construct the dissimilarity measure between any two markets. To demonstrate the dynamics of the changes, three sub-periods are considered: the period before the global financial crisis (from October 2002 to July 2007), the period of the crisis itself (from July 2007 to December 2008) and the post-crisis period (from January 2009 to April 2012). Taking dynamical relationships into account, all stock markets can be divided into four clusters: North and South America, Western Europe, Eastern Europe and Asia. However, in each of these main clusters similarities between financial markets vary with time.

1. Introduction

The analysis of similarities between financial markets all over the world is important for the investigations. Practitioners are interested in identifying these similarities for investment and risk management purposes. This knowledge about market relationships enables them to diversify the risk of their investments. Furthermore, of the same importance is knowledge about clustering financial time series in order to understand the internal relationship of financial markets. This analysis, which consists in creating a set of markets according to a certain measure of similarity of their behavior, has appeared to be very helpful in formulating investment strategies. However, this procedure involves the construction of a convenient similarity measure of the empirical data that would take into account the characteristics of financial time series.

Mantegna (1999), Bonanno, Lillo and Mantegna(2001), among others, used a Pearson correlation coefficient to construct the dissimilarity measure of a pair of stocks. However, it is known that in the case of financial time series clustering this measure does not use information about the autocorrelation structure of each stock return and it does not take into account information about the return volatility. This approach is only suitable for multivariate elliptical distributions (Musetti, 2012). Piccolo (1990) introduced a dissimilarity measure for an autoregressive process. This measure was extended to the GARCH process family by Otranto (2004). In this

* We would like to thank two anonymous referees for the valuable comments on an earlier version of the paper.

work, daily returns of nine stock exchange indices from 1995 to 2001 were considered. In this case, proximity is identified if the underlying conditional variance equations are similar in terms of their structure. This principle, however, seems to be less accurate in the case of analyzing market returns due to the fact that the estimates of their parameters to be compared do not usually vary materially from one another.

The other approach to cluster GARCH processes was discussed by Caiado and Crato (2007), who introduced an approach based on information about the estimated GARCH parameters for the considered financial time series. This procedure was used for clustering 27 international stock markets using daily return series with different sample sizes from 1966 to 2006. The data were divided into two sample periods: prior and subsequent to the terrorist attacks on September 11, 2001. The main conclusion was that the majority of European markets, the United States and Canada appeared close to one another, and most Asian/Pacific markets and the South/Central American markets formed a separate cluster. Since the terrorist attacks, the European stock markets have become more homogenous, and the North American markets, Japan and Australia seem to have become closer.

Clustering of international stock markets was presented also in the paper by Bastos and Caiado (2009). This approach was based on variance ratio test statistics. The authors divided the sample into three sub-periods and found that the markets generally connected according to their size and level of development, although in the clusters there are markets from different geographical regions.

From the literature review it seems obvious that market clustering differs in sub-periods. This fact requires an approach that would take into account not only that clustering could change in sub-periods but also allows for presentation of fully dynamic market clustering. For the purposes of constructing a dynamic measure of dissimilarity between markets, it is useful to study their dynamical dependencies.

In the literature there have been proposed models in which dynamic correlations between time series have been considered, such as the dynamic conditional correlation model of Engle (2002) or the regime switching dynamic correlation model proposed by Pelletier (2006). However, when the number of considered time series is large, there is a problem with parameter estimation. Furthermore, the asymmetry inherent in conditional distribution is an additional problem in the case of parameter estimation. We try to solve this problem using the approach based on the copula function theory, where the dependence structure and the margins can be separated in multivariate continuous distribution functions. The time-varying dependencies between time series may be explained by the regime switching process. This approach has been widely discussed in the literature. For example, Patton (2006 and 2009) studied time-varying copulas for modelling asymmetric exchange rate dependence; Bartram et al. (2007), following the methodology of Patton (2006) and using a conditional Gaussian copula for this purpose, estimated time-varying copula dependence models for 17 European stock market indices. Okimoto (2008) estimated regime switching copulas for the US-UK pair. Kenourgios, Samitas and Paltalidis (2011) used both a multivariate regime switching Gaussian copula model and the asymmetric generalized dynamic conditional correlation approach to capture non-linear correlation dynamics during five recent financial crises. The regime switching copula model with a Markov switching mechanism for modeling financial time series has been discussed also by Jondeau and Rockinger (2006), Rodriguez (2007), Chollete at al. (2009) and others.

Numerous studies on dependence dynamics among the markets have been presented. In particular, there have been many attempts to examine market relationships before and after crises have, though there are still few studies exploring the consequences of such co-movement for market clustering. Our investigation builds on the foundation of extensive research that has dealt with the subject of dynamic correlation between financial markets by applying the dynamic framework to the notion of clustering. To the best of our knowledge it is the first investigation of this kind.

In this work we discuss the Markov switching copula model employed to describe the dynamics of the relationships between considered time series. The time series under consideration are characterized by high kurtosis and skewness and for this reason the copula-based approach is used in this study. We define the conditional Spearman coefficient as a measure of similarity between each two pairs of time series. The matrix of these parameters denoted by each t is a base for construction of the dissimilarity measure for clustering purposes. This approach allows clustering of markets at any time t and forecast. The study covers daily returns of 36 market indices from all over the world. It is a random sample of markets, diversified enough to capture all specific properties in terms of both geographical and economic dimensions. It includes markets that are deemed to be emerging as well as the already developed markets of North and South America, Europe and Asia. Daily returns come from the period from October 2002 to April 2012. For researching the dependencies between these data the regime switching copula models with two regimes is considered. The Markov switching model in a pair composed of Normal, t-Student or Joe-Clayton copulas is exploited. The choice of the copulas is based on their different tail dependence properties.

In order to show the dynamic of changes in clustering, the analysis is carried out for each t. It is impossible to show results for each t; therefore the cluster analysis results are presented in some sub-periods. The selection of the periods is conditioned by the behavior of the US market. Firstly, the entire sample, i.e. from October 2002 to April 2012, is considered to present the general similarity of markets. Next the following sub-periods are analyzed: the period before the global financial crisis (from the beginning of the sample to July 2007), the period of the crisis itself (from July 2007 to December 2008) and finally, the post-crisis period (from January 2009 to the end of the sample, i.e. April 2012). These periods are specified to show the dynamics of the changes taking place in the studied markets.

The analysis results of the study show that on average four major clusters are generally observed: the American markets, the West European markets, the Eastern European markets and the Asian markets. The clustering changes that occur over time take place only within these major clusters.

The main body of the paper is organized as follows: In Section 2 the model definition is presented. The empirical study is presented in Section 3 with three subsections: the studied data are described in Section 3.1, some results obtained from the estimation of the regime switching copula models with two regimes are discussed in Section 3.2 and finally, the clustering results are shown in Section 3.3.

2. Model

2.1 Distributions of Returns

In the case of financial time series modeling the GARCH(1,1) model proposed by Bollerslev (1986) is the simplest and the most popular parameterization. However, limiting the number of lags in the variance equation to exactly one for the GARCH effect is not always justified by the data. Thus, two additional properties of the returns need to be considered. The first of them is associated with the autocorrelation of time series. At the same time, the autocorrelation of stock returns vanishes very rapidly for higher lags, so that it is sufficient in most practical applications to include only one autocorrelation term. The second property is that the effect of positive and negative returns on the variances differs in terms of its magnitude. This leads to the AR(1)-GJR-GARCH(1,1) model (Glosten, Jagannathan and Runkle, 1993), where the parameterization is as follows:

$$
\begin{aligned}
& y_t = \mu_t + \varepsilon_t \,,\ \mu_t = \mu + \phi y_{t-1} \,,\ |\phi| < 1 \\
& \varepsilon_t = \sqrt{h_t}\,\xi_t \,,\ \xi_t \sim IID(0,1) \\
& h_t = \kappa + \alpha \varepsilon_{t-1}^2 + \alpha^- \varepsilon_{t-1}^2 I_{t-1} + \beta h_{t-1}
\end{aligned}
\tag{1}
$$

The last equation presented above contains a binary variable responsible for capturing the above-mentioned asymmetry effect:

$$
I_t = \begin{cases} 1 & \text{if } \varepsilon_t < 0 \\ 0 & \text{otherwise} \end{cases}
$$

Scrutiny of daily returns led to the introduction of fat-tailed with high kurtosis and skewed distributions of the residuals ξ_t. Hansen (1994) proposed a variant of the GARCH model in which the first four moments are conditional and time varying. In this model the conditional distribution function is skewed *t*-Student with parameters η_t and λ_t denoting the degree of freedom and the asymmetry parameter, where

$$
\begin{aligned}
& \eta_t = \omega_\eta + \theta_\eta \varepsilon_{t-1} + \theta_\eta^- \varepsilon_{t-1} I_{t-1} + \rho_\eta \eta_{t-1} \\
& \lambda_t = \omega_\lambda + \theta_\lambda \varepsilon_{t-1} + \theta_\lambda^- \varepsilon_{t-1} I_{t-1} + \rho_\lambda \eta_{t-1}
\end{aligned}
\tag{2}
$$

A similar margin model was considered in the work of Jondeau and Rockinger (2006) for modeling four major stock indices. From their empirical studies it can be concluded that this model is quite satisfactory for this kind of financial time series.

2.2 Regime Switching Copula Model

The regime switching model is used to model the dependence of market index returns. It is natural to consider the two regimes characterized by different shapes of dependencies. The Markov switching copula model is considered, whereas the binary variable indicating the current regime is assumed to be latent and the marginal distributions are not dependent on the regime. The parameter of the model is obtained using the maximum likelihood method. Because the research involves a large number of markets, the bivariate model is discussed instead of considering the multivariate regime switching copula already presented by Chollete at al. (2009).

The regime is indexed by an unobserved random variable S_t. Let x_t, $t = 1,...,T$ be the two-dimensional vector of observed variables and let Ω_T be a vector containing all observations obtained through date t. The density of x_t conditional on S_t taking on the value j is:

$$f\left(x_t \middle| S_t = j, \Omega_{t-1}\right) =$$
$$= c^{(j)}\left(F_{1,t}\left(x_{1,t} \mid S_t = j, \Omega_{t-1}\right), F_{2,t}\left(x_{2,t} \mid S_t = j, \Omega_{t-1}\right); \theta_c^j\right)\prod_{i=1}^{2} f_i\left(x_{i,t}, \theta_m^i\right) \quad (3)$$

where $c^{(j)}$ denotes a copula density with parameter θ_c^j being in the regime j and $f_i\left(x_{i,t}; \theta_m^i\right)$ is the density of the marginal distribution with parameters θ_m^i, $i = 1, 2$. The unobserved latent state variable follows a Markov chain with the transition matrix:

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

where $p_{11} = P\left(S_t = 1 \middle| S_{t-1} = 1\right)$, $p_{22} = P\left(S_t = 2 \middle| S_{t-1} = 2\right)$, $p_{12} = 1 - p_{11}, p_{21} = 1 - p_{22}$.

Let θ be a vector of all unknown parameters. The log likelihood function for the observed data Ω_T has the form:

$$L(\theta) = \sum_{t=1}^{T}\sum_{j=1}^{2} \log\left(f\left(x_t \middle| S_t = j, \Omega_{t-1}\right) P\left(S_t = j \middle| \Omega_{t-1}\right)\right) = L_c(\theta) + L_m(\theta) \quad (4)$$

where

$$L_c(\theta) = \sum_{t=1}^{T}\sum_{j=1}^{2} \log\left(c^{(j)}\left(F_{1,t}\left(x_{1,t} \mid S_t = j, \Omega_{t-1}\right), F_{2,t}\left(x_{2,t} \mid S_t = j, \Omega_{t-1}\right); \theta_c^j\right) P\left(S_t = j \middle| \Omega_{t-1}\right)\right) \quad (5)$$

and

$$L_m(\theta) = \sum_{t=1}^{T}\sum_{i=1}^{2} \log f_i\left(x_{i,t}, \theta_m^i\right) P\left(S_t = j \middle| \Omega_{t-1}\right) \quad (6)$$

As we mentioned before, the marginal densities are not regime-dependent, which is why the estimation of the margins do not depend on the history. The procedure was performed in two steps. Firstly, the margin distribution parameters were estimated. Secondly, the estimation of regime switching copula model, which is conditional on having estimated the marginal models, is carried on. Given the fact that the Markov chain S_t is not observable, the Hamilton filter (for details, see Hamilton, 1994) is used for estimation purposes. For simplicity, let $\eta_{j,t}$ denote the copula densities under the j regime:

$$\eta_{j,t} = c^{(j)}\left(F_{1,t}\left(x_{1,t} \middle| \Omega_{t-1}\right), F_{2,t}\left(x_{2,t} \middle| \Omega_{t-1}\right); \theta_c^j\right) \quad (7)$$

and $\hat{\xi}_{j,t|t-1}$ denotes the probability

$$\hat{\xi}_{j,t|t-1} = P\left(S_t = j \middle| \Omega_{t-1}\right) \quad (8)$$

Therefore, the aim is finding the maximum of:

$$L_c\left(\theta|\hat{\theta}_m^i\right)=\sum_{t=1}^{T}\log\sum_{j=1}^{2}\eta_{j,t}\,\hat{\xi}_{j,t|t-1} \tag{9}$$

For each i $j\left(j=1,2\right)$, for t, using the Hamilton filter we have:

$$\hat{\xi}_{j,t|t-1}=\sum_{i=1}^{2}p_{ij}\hat{\xi}_{i,t-1|t-1} \tag{10}$$

where

$$\hat{\xi}_{i,t-1|t-1}=P\left(S_{t-1}=i\,|\,\Omega_{t-1}\right) \tag{11}$$

Generally:

$$\hat{\xi}_{j,t|t}=P\left(S_t=j\,|\,\Omega_t\right)=\frac{\eta_{j,t}\hat{\xi}_{j,t|t-1}}{\sum_{j=1}^{2}\eta_{j,t}\hat{\xi}_{j,t|t-1}} \tag{12}$$

The algorithm of the likelihood function construction is performed iteratively for $t=1,\ldots,T$. In each t step we accepted the values $\hat{\xi}_{j,t-1|t-1}$ as an input and next we calculated the component of $L_c\left(\theta|\hat{\theta}_m^i\right)$ and obtained the output $\hat{\xi}_{j,t|t}$.

With the filter described above, we obtain the probability distribution of S_t given the information set by t. However, it is useful to know the distribution of S_t given the full sample information set, i.e. all T observations. The smoothed probabilities $\hat{\xi}_{j,t|T}=P\left(S_t=j\,|\,\Omega_T\right)$ may be obtained according to the procedure described in Hamilton(1994).

2.3 Spearman Measure of Association

The population version of the Spearman coefficient is defined to be proportional to the probability of concordance minus the probability of discordance (Nelsen, 1999) for a pair of vectors with the same margins, but one vector has the distribution function F, while the components of the other are independent. It can be shown (Nelsen, 1999, Theorem 5.1.6) that this ratio has the form:

$$\rho=12\iint_{0}^{1}C\left(u\right)du-3 \tag{13}$$

Let ρ_j (j = 1,2) be a Spearman coefficient connected with a copula function $C^{(j)}$ in the regime j and let ρ_t be a Spearman coefficient at time t. It is quite obvious that due to the fact that we are dealing with the two regimes, this ratio should also be dependent on the time. The expected value of the Spearman correlation at time t is:

$$E\left(\rho_t\right)=\sum_{j=1}^{2}\rho_j P\left(S_t=j\right) \tag{14}$$

The expected value of the Spearman correlation, conditional on Ω_T, at time t, has the form:

Table 1 Descriptive Statistics of the Data

	Average	Median	Std dev.	Skewness	Kurtosis
Min	-0.026	0.034	0.945	-0.927	2.667
Average	0.044	0.117	1.728	-0.337	7.349
Max	0.106	0.232	2.467	1.031	23.160
Std dev.	0.029	0.054	0.335	0.375	3.848

$$E\left(\rho_t \mid \Omega_T\right) = \sum_{j=1}^{2} \rho_j P\left(S_t = j \mid \Omega_T\right) \tag{15}$$

Having a full sample information set, Ω_T, estimates $\hat{\xi}_{j,t|T}$ and $\hat{\rho}_j$, it is possible to estimate the conditional expected value of the Spearman correlation, $\hat{\rho}_{t|T}$, at time t as:

$$\hat{\rho}_{t|T} = \sum_{j=1}^{2} \hat{\rho}_j \hat{\xi}_{j,t|T} \tag{16}$$

The forecasting coefficient at time ($T + 1$) may be calculated as:

$$\hat{\rho} = \sum_{j=1}^{2} \hat{\rho}_j \hat{\xi}_{j,T+1|T} \tag{17}$$

3. Empirical Results

3.1 Data

The investigation covers returns of market indices from all over the world. It is a random sample, diversified enough to capture all specific properties in terms of both geographical and economic dimensions. It includes markets that are deemed to be emerging as well as the already developed markets of North and South America, Europe and Asia. In the case of the United States, two different indices, the Dow Jones Industrial Average and NASDAQ, are included because of the enormous impact of the US market on the finances of the rest of the world. In addition, the BEL20, the benchmark stock market index of Euronext Brussels, is also included in the research. The indices are all denominated in US dollars because the daily data for stock markets of different regions are considered. Daily returns are from the period from October 2002 to April 2012. To deal with the missing data in the sample, the linear approximation is used. The daily returns are computed as the difference between the logarithm of price on day t and the logarithm of price on day $t - 1$. *Table 1* contains the descriptive statistics of these data.

One can see from *Table 1* that the average of the returns ranges from –0.026% to 0.106%. Generally, the median of the considered return is greater than its average, whereas the skewness is negative, so the skewed distribution of returns should be taken into consideration. The kurtosis takes values from 2.667 to 23.160, so a tendency toward a high kurtosis should also be taken into account in the modeling of returns. In order to examine the properties of the time series, especially autocorrelation and heteroscedasticity, the Ljung-Box and Engle tests are carried out.

The test results indicate the existence of a first lag autocorrelation and GARCH effect for almost all considered returns of the indices.

3.2 Model Estimation Results

In the preliminary step of our empirical work, we investigate the structure of the univariate marginal returns. The model suggested for the description of the returns is based on the autocorrelation and heteroskedasticity test results. The effect of positive and negative returns on the variance is also included in the model. First the model AR(1)-GJR-GARCH(1, 1) with Hansen's skewed t-Student conditional distribution is considered to describe the modeling of returns. Thus, the procedure of testing the goodness-of-fit is carried out. For testing purposes, we follow the procedure described in Diebold et al. (1998). If a marginal distribution is correctly specified, the margins denoting the transformed standardized AR(1)-GJR-GARCH(1,1) residuals should be iid Uniform (0,1). In the cases when test results are doubtful the time-varying volatility, skewness and kurtosis are taken into consideration in Hansen's conditional distributions in order to improve tests results. A similar procedure was used in the work of Jondeauand Rockinger (2006). Most of the considered time series test results confirm the correctness of the chosen model. The aim of applying this procedure is to improve the quality of fitting the model to the data in order to obtain the uniform distribution necessary to carry out the estimation of the Markov switching copula model.

It is impossible to estimate the d-multidimensional Markov switching copula model parameters for $d = 36$ due to the fact that the number of variables is too big. Thus, the bivariate model for all market pairs is estimated. Three copulas, namely the normal copula, t-Student copula and Joe-Clayton copula, are taken into consideration and the four regime switching models—normal-normal, normal-Joe-Clayton, t-Student-Joe-Clayton and t-Student-t-Student—are obtained. The statistical significance of the parameters is analyzed and goodness-of-fit tests for these models are carried out. The applied goodness-of-fit tests are based on the reasoning of Breymann et al. (2003) and of Genest, Remillard and Beaudoin (2009). The percentage of cases where there is no reason to reject the null hypothesis, presuming that a given pair of series can be modeled with the assumed regime switching model, is close to 87% for the t-Student-t-Student copula and close to 86% for the t-Student-Joe-Clayton copula at the significance level $\alpha = 0.05$ and it is close to 95% for both models at the significance level $\alpha = 0.01$. It is worth noting that these models seemed to surpass the others, especially for the normal-normal model. This conclusion is consistent with our expectations and is additionally confirmed by a comparison of the AIC values.

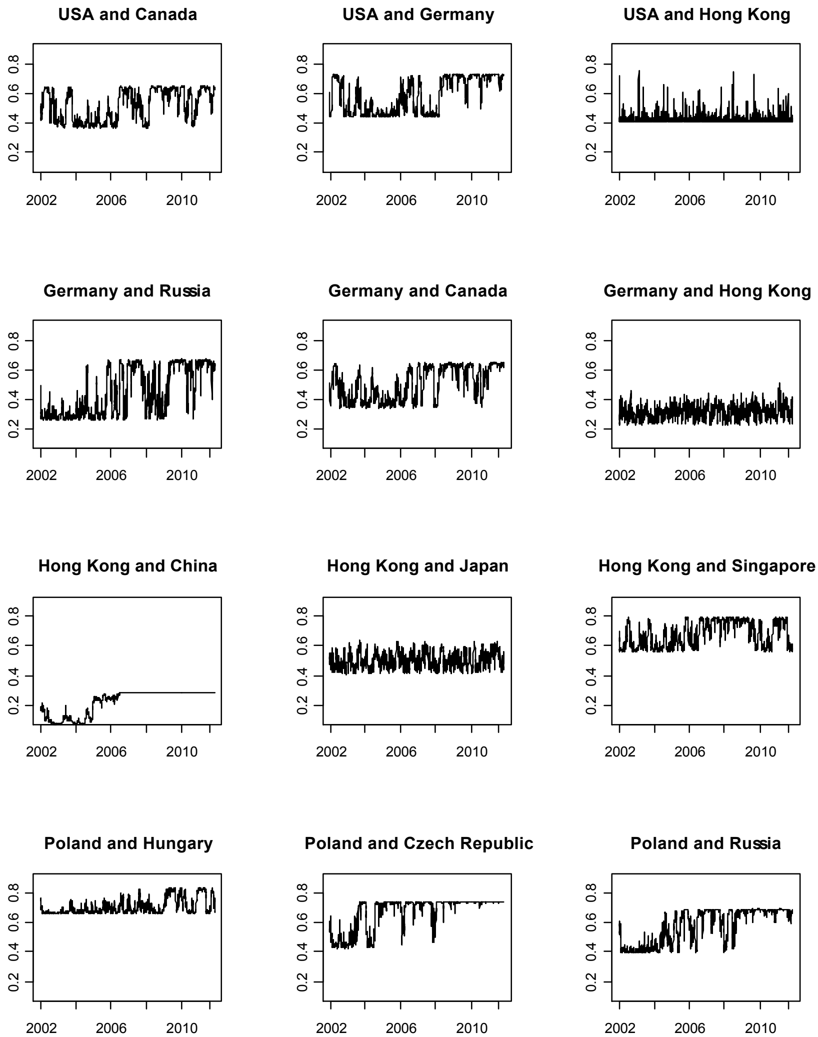
Analyses of the significance of the Markov switching model parameters reveal the presence of a high correlation regime and a low correlation regime in more than 50% of all considered cases. The presence of the two regimes is independent from the considered model. Conversely, the insignificant regime of one model is also accompanied by the insignificant regime of the other. So, taking into account the above argumentation, the results of the t-Student-t-Student regime switching model estimation may be appropriate for drawing conclusions on the behavior of dependencies between markets. Let $\hat{\rho}_1$ be an estimate of the Spearman correlation parameter for the low correlation regime and let $\hat{\rho}_2$ be an estimate of the Spearman correlation parameter for the high correlation regime. These parameters are calculated numerically. The average and the standard deviation of $\hat{\rho}_1$ are 0.27 and 0.17, respectively, whereas the average and the standard deviation of $\hat{\rho}_2$ are 0.57 and 0.19, respectively. The strongest connections are found between the countries in Western Europe. The two regimes are observable and both of them indicate strong interdependence between these countries ($\hat{\rho}_1$ is close to 0.83 on average and $\hat{\rho}_2$ is close to 0.95 on average). The two regimes are present for the countries in other regions as well. However, in none of them are the results of the two regime correlations as high as for Western Europe. For example, for the countries in Eastern Europe—Poland, Hungary, the Czech Republic and Russia—parameters $\hat{\rho}_1$ and $\hat{\rho}_2$ range from 0.18 to 0.66 in the lower dependency regime and from 0.56 to 0.83 in the higher dependency regime. For the studied countries of Asia the average and the standard deviations of parameter $\hat{\rho}_1$ are 0.26 and 0.15, whereas the average and the standard deviation of parameter $\hat{\rho}_2$ are 0.58 and 0.19. For the North and South American countries, the received average and standard deviation are 0.34 and 0.08 for $\hat{\rho}_1$ and 0.65 and 0.09 for $\hat{\rho}_2$.

Thus, for each pair (i, j) of the indices the time-varying weighted Spearman correlation coefficient $\hat{\rho}_t^{ij}$ is calculated using formula (16). Some results are presented in *Figure 1*. *Figure 1* shows the interdependence between the US market (DJIA index) and the neighboring Canadian market, the interdependence between the US market and the index of Germany, which represents the European Union, and the interdependence between the US market and a selected Asian market (Hong Kong). Because Hong Kong is one of the world's most important financial centers, it was selected for this analysis (in the case of the analysis of the last pair of indices, the time-zone difference is taken into account). Between the US index and the Canadian or German indices, two levels of dependencies are clearly visible: low interdependence (approximately 0.4) and strong interdependence (approximately 0.65). The power of dependencies varies over time and it is similar in both analyzed pairs. The interdependence between the US and Hong Kong markets is moderate and remains at the same level throughout the study.

Next, *Figure 1* presents the time-varying interdependence between the German market and the Russia market, where two levels of interdependence are observed—weak (approximately 0.25) and moderate (approximately 0.60), the also two-level (0.38 and 0.60) time-varying interdependence between the German market and the Canadian market, and the rather stable time-varying interdependence between the German market and the Hong Kong market with only one prevailing regime. It can be noticed that the interdependence between the Hong Kong and German markets is a little weaker than the relationship between the Hong Kong and US markets.

Furthermore, *Figure 1* presents the results of the interdependence between the selected Asian markets. Three pairs of markets: Hong Kong-China, Hong Kong-Japan and Hong Kong-Singapore were included in the analysis. The interdependence between the Hong Kong and Chinese markets is relatively weak (alternating from

Figure 1 Time-Varying Spearman Correlations Between Some Markets



0.05 to 0.3). The interdependence between the markets of Hong Kong and Japan is almost homogeneously moderate, as the time-varying Spearman correlation oscillates at the level of 0.4 to 0.6. In the case study of the dependence between the Hong Kong and Singaporean markets, we can clearly identify two levels sustained over a long period of time: one moderate (approximately 0.5) and the other strong (approximately 0.8). Finally, we present the relationship between the Polish market and the markets of Hungary, the Czech Republic and Russia. These markets represent Eastern Europe. In the studied period we can see that the Polish market always correlates with the Hungarian market and adopts two levels of correlations that are quite high, from 0.60 to 0.85. The analysis of the correlations between the Polish and

Czech and between the Polish and Russian markets indicate time-varying interdependence, with the levels ranging from 0.42 to 0.75 and from 0.4 to 0.7, respectively. It can be noted that for both pairs the interdependence of these markets increases with the passage of time. In most cases in which the two regimes exist, the interdependency has a tendency to increase with time. In the majority of the pairs, the transition to a higher level of interdependency coincided with the occurrence of the crisis in financial markets, which may suggest that during a crisis the correlations between the markets are stronger than in times of prosperity. However, this observation may be the result of a general tendency to globalize not only financial markets but also national economies.

3.3 Cluster Analysis

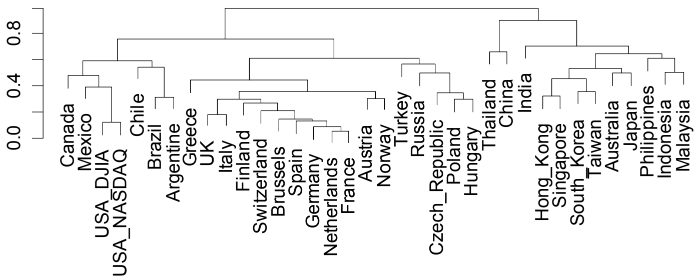
From the analysis carried out above it can be concluded that the markets are generally connected to one another, with a few exceptions. Therefore, it would be interesting to indicate which of them have some similarities in their dependencies. Furthermore, the results of the above analysis indicate the fact that the intensity of dependencies between financial markets changes over time. The change in intensity of the dependenceis had an impact on the dynamics of market clustering. We define a dissimilarity matrix, $D_t = \left(d_t^{ij}\right)_{i,j=1\ldots,N}$ where $d_t^{ij} = 1 - \hat{\rho}_t^{ij}$, to cluster financial data for each t.

The clustering algorithm used in this empirical work is based on the Wards algorithm. It is impossible to show results for each t, so the cluster analysis is performed in some sub-periods. In this case $D = \left(d^{ij}\right)_{i,j=1\ldots,N}$ where $d^{ij} = \frac{1}{T_0}\sum_{t=1}^{T_0} d_t^{ij}$ is taken into consideration as a dissimilarity matrix. It should be noted that the time-varying Spearman correlation coefficient $\hat{\rho}_t^{ij}$ is estimated for the entire sample.

The selection of the periods is conditioned by the behavior of the US market. So, four sub-periods were considered: the entire sample from October 2002 to April 2012, the period before the global financial crisis (from beginning of the sample to July 2007), the period of the crisis itself (from July 2007 to December 2008) and, finally, the post-crisis period (from January 2009 to the end of the sample, i.e. April 2012). Time-zone differences are taken into account in the empirical study. Since the daily data of stock markets in different regions are used, a problem with non-synchronous trading that can cause bias in the estimation appeared. In order to take this fact into account, lagged American returns are used, although the trading times partially overlap and the Asian returns are taken from the following day. However, this procedure results in a difference of two days between American and Asian trading times. So, the case where only Asian returns are taken from the following day and American returns are taken from the same day as European returns is also discussed. However, the use of both procedures does not have a major impact on the clustering results.

When the entire sample is taken into consideration (*Figure 2*), the following clusters are clearly observed: i) the American markets, ii) the European markets and iii) the Asian markets. Those clusters are transparent and the result corresponds with our prior expectations.

Figure 2 Dendrogram for the Index Returns of the American, Asian and European Markets for the Entire Research Sample



In the first cluster, two subgroups are clearly noticeable: the subgroup of North American indices and the subgroup of South American indices. Both US indices are highly correlated and the Mexican and Canadian indices are linked to them. The Brazilian market is more closely related to the Argentinean market than to the Chilean market and these three indices form the subgroup of South American indices.

In the second cluster, the Western European indices (France, the Netherlands, Germany, Spain, Switzerland, Finland, UK and Italy) are strongly correlated with each other and they are classified into one subgroup. The Austrian and Norwegian markets are in a separate subgroup and the Greek market seems to be isolated from the other markets of the Western European cluster. The Polish and Hungarian markets tend to be the most similar to each other in the group of Eastern European markets, and they are less similar by degrees to the Czech, Russian and Turkish markets.

In the cluster of the Asian markets, the strongest connections can be noticed between the Hong Kong and Singaporean markets and the South Korean and Taiwanese markets, which comprise one subgroup. The Australian and and Japanese markets comprise a separate subgroup together with Indonesia, Malaysia and Philippines. A distinct subgroup of the Asian indices is formed by Thailand and China. The Indian market seems to be isolated from other markets. Our findings are generally in line with other studies that use different measures of dissimilarity (among others Caiado, Crato, 2007).

To demonstrate the dynamics of changes in the market clustering, we present the results in consecutive sub-periods. When the period before the global crisis is analyzed, *Figure 3a*, the index clusters are similar to those discussed above. In the cluster of Asian markets, some markets changed their position compared with that previously considered. The Chinese market is isolated and not correlated with any other of the researched markets.

During the crisis period (*Figure 3b*), there are minor differences in the grouped indices in comparison with the previous results. For example, in the cluster of Western European indices, the position of the Finnish market is changed. It seems to be less connected with the euro-zone markets and it is more closely linked to the Austrian market, with which it forms a separated subgroup. Furthermore, the Russian market is less closely related to others, which is the opposite of

Figure 3a Dendrogram for the Index Returns of the American, Asian and European Markets for the Period Before the Global Crisis (from the Beginning of the Sample to July 2007)

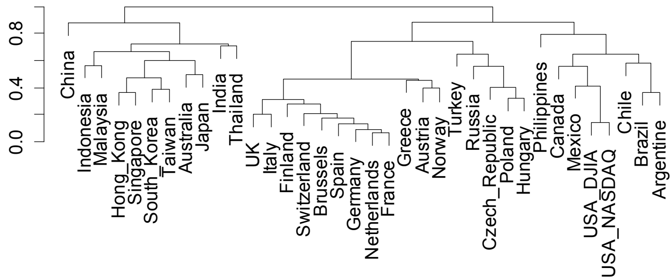


Figure 3b The Dendrogram for the Index Returns of the American, Asian and European Markets for the Period of the Global Financial Crisis

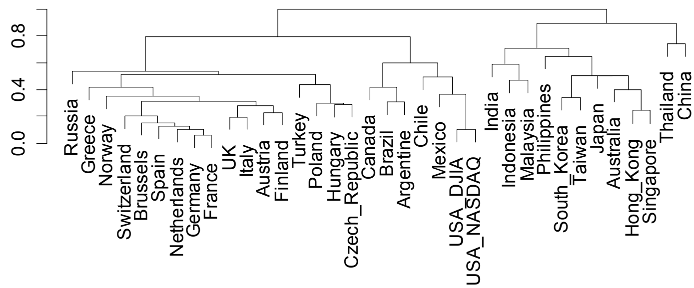
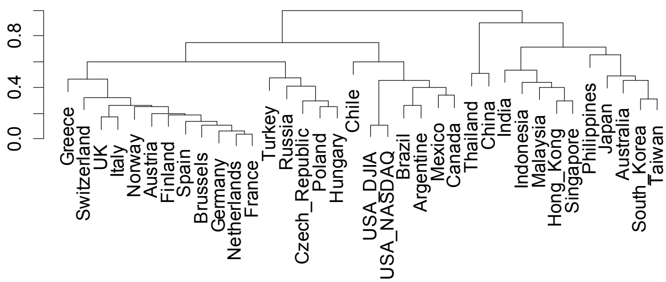


Figure 3c The Dendrogram for the Index Returns of the American, Asian and European Markets for the Post-Crisis Period



the previously considered case. Slight differences are also noticeable in the cluster of American indices, which is caused by the repositioning of the Canadian and Chilean indices. The differences are also noticeable inside the cluster of Asian markets. The Thai market connects with the Chinese market, whereas the Indian market becomes separated.

Further differences were observed during the last period studied (*Figure 3c*). The relationships within the American markets as well as within the Eastern

European markets are similar to the corresponding ones from the period before the crisis. However, the US indices seem to be more separated. The markets in Western Europe are more integrated than in the both previous periods and all of them come into in one set. In the cluster of Asian markets, small differences, when compared with the previous period, seem insignificant. Hong Kong's and Singapore's markets changed their subgroup and they are now in the set with the markets of Indonesia, Malaysia and India.

However, the fact that we clearly defined major clusters—American markets, Western European markets, Eastern European markets and Asian markets—is apparent in the clustering studies in sub-periods. The clustering changes that occur over time take place within these major clusters. The crisis did not significantly affect the results of clustering. The changes that have occurred in the market groupings may be due to the internal economic situation in a given country.

4. Conclusions

The aim of this work was to attempt dynamical clustering of index returns of European, American and Asian stock markets. The approach taken allows not only describing the dynamical structure of time series but also offers a way to describe dynamic cluster historical data and a way to predict clustering in the next time period. In order to examine the dependencies between these data, regime switching copula models with two regimes were considered. An analysis was carried out for all pairs created from the considered indices. In most of the pairs where two regimes existed, one can notice that dependency has a tendency to increase with time. The transition to a higher level of interdependency coincided with the occurrence of the crisis in financial markets, which may suggest that during a crisis the correlations between the markets are stronger than in times of prosperity. However, this observation may be a result of a global tendency to globalize not only financial markets but also national economies.

The change in intensity of the dependences influenced the dynamics of market clustering. The time-varying correlation coefficient obtained from the Markov switching copula model is suitable for deriving a dissimilarity measure between the time series under consideration for each time t. Since it was impossible to describe results for each t, the cluster analysis was performed in some sub-periods. To demonstrate the dynamics of change in the similarity between markets, some sub-periods were considered: the period before the global financial crisis, the period of the crisis and the post-crisis period. Taking into account the dynamical relationships, all stock markets can be divided into four clusters: North and South America, Western Europe, Eastern Europe and Asia. Generally, the clustering changes that occur over time take place only within these major clusters. The fact what the markets were under stress (due to the crisis) did not significantly affect the results of clustering. The changes that have occurred in the market groupings may be due to the internal economic situation in a given country. It seems, therefore, that in order to diversify risk one should invest in markets that differ geographically.

REFERENCES

Bartram SM, Taylor SJ, Wang YH (2007): The Euro and European financial market dependence. *Journal of Banking & Finance*, 31:1461–81.

Bastos JA, Caiado J (2009): Clustering global equity markets with variance ratio tests. *CEMAPRE Working paper*, no. 0904.

Bollerslev T (1986): Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307-327.

Bonanno G, Lillo F, Mantegna R (2001): Level of complexity in financial markets. *Physica A*, 299: 16–27.

Breymann W, Dias A, Embrechts P (2003): Dependence Structures for Multivariate High-Frequency Data in Finance. *Quantitative Finance*, 3:1–14.

Caiado J, Crato N (2007): A GARCH-based method for clustering of financial time series: International stock markets evidence. *MPRA Paper,* no. 2074.

Chollete L, Heinen A, Valdesogo A (2009): Modeling international financial returns with a multivariate regime switching copula. *Journal of Financial Econometrics*, 7(4):437–80.

Diebold FX, Gunther TA, Tay AS (1998): Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4):863–83.

Engle RF (2002): Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20:339–50.

Genest R, Remillard B, Beaudoin D (2009): Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–214.

Glosten L, Jagannathan R, Runkle D (1993): On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48:1779–1801.

Hamilton J (1994): *Time series analysis*. Princeton Univ. Press.

Hansen B (1994): Autoregressive Conditional Density Estimation. *International Economic Review*, 35:705–730.

Jondeau E, Rockinger M (2006): The Copula-GARCH model of conditional dependencies: An international stock market application. *Journal of International Money and Finance*, 25:827–53.

Kenourgios D, Samitas A, Paltalidis N (2011): Financial crises and stock market contagion in a multivariate time-varying asymmetric framework. *Journal of International Financial Markets, Institutions & Money*, 21(1):92–106.

Mantegna RN (1999): Hierarchical structure in financial markets. The *European Physical Journal B*, 11:193–7.

Mirkin B (2005): *Clustering for Data Mining: A Data Recovery Approach*. Boca Raton (Fl.), Chapman and Hall/CRC.

Nelsen B R (1999): *An Introduction to Copulas*. Springer Verlag, New York.

Okimoto T (2008): New evidence of asymmetric dependence structures in international equity markets. *Journal of Financial and Quantitative Analysis*, 43:787–815.

Otranto E (2004): Classifying the Markets Volatility with ARMA Distance Measures. *Quaderni di Statistica*, 6:1–19.

Patton AJ (2006):Modelling asymmetric exchange rate dependence. *International Economic Review*, 47:527–56.

Patton AJ (2009):Copula-based Models for Financial Time Series. *Handbook of financial time series*.

Pelletier D (2006): Regime-switching for dynamic correlation. *Journal of Econometrics*, 131: 445–73.

Piccolo D (1990): A distance measure for classifying ARIMA models. *Journal of Time Series Analysis*, 11:153–64.

Rodriguez JC (2007): Measuring Financial Contagion: A Copula Approach. *Journal of Empirical Finance*, 14(3):401–23.